

UNIVERSIDAD COMPLUTENSE DE MADRID
FACULTAD DE FILOSOFÍA
Departamento de Filosofía I
(Metafísica y Teoría del conocimiento)



LAS CIENCIAS COGNITIVAS Y SU FUNDAMENTACIÓN
FILOSÓFICA

MEMORIA PARA OPTAR AL GRADO DE DOCTOR
PRESENTADA POR

Óscar Lucas González Castán

Bajo la dirección del doctor:
Miguel García-Baro López

Madrid, 2002

• ISBN: 978-84-8466-255-6

© Óscar Lucas González Castán, 1993

UNIVERSIDAD COMPLUTENSE

Facultad de Filosofía

**LAS CIENCIAS COGNITIVAS
Y SU FUNDAMENTACION FILOSOFICA:
SIGNIFICADO Y CONCIENCIA**

OSCAR L. GONZALEZ-CASTAN.

Tesis Doctoral dirigida por el Dr. D. Miguel García-Baró,
Profesor Titular del Departamento de Filosofía I
de la Universidad Complutense.

MADRID, 1993

D. AUA. N.º R. I. O. J. A. ... M. E. I. O.
 Secretario del Tribunal calificador de la
 Tesis Doctoral de D. OSCAR L. GONZALEZ
 ... GASTAN. dirigida por el
 Dr. D. MIGUEL G.º - BARD. ... L. O. P. E. Z.
 CERTIFICO. ... Presente Tesis ha sido
 defendida ... el ... de la Fecha ante el
 Tribunal nombrado al efecto, haciendo ob-
 tenido la calificación de Apta. Cum.
Valde por unanimidad
 ... en Madrid a ... 6. de ...
 ... Juris. de ... 1993.

V.º R.º
 EL PTE. DEL
 TRIBUNAL

EL SECRETARIO

INDICE GENERAL

PRIMERA PARTE.

EN TORNO AL SER PSIQUICO: SENTIDO COMUN, CIENCIA Y METACIENCIA.

CAPITULO 1.

SENTIDO DE ESTAS INVESTIGACIONES.

& 1. El problema de Sócrates	2
& 2. Sócrates y la caracterización del ser psíquico	4
(a) Imprescindibilidad y profundidad de la psicología popular	4
(b) El vocabulario intencional	9
(c) Las cláusulas <i>ceteris paribus</i>	14
(d) Intencionalidad y causalidad	17
(e) La promiscuidad ontológica del ser psíquico	21
& 3. Husserl y la caracterización del ser psíquico	27
(a) Intencionalidad e "imperar" yoico	27
(b) Las cinestesias: ¿La glándula pineal de Husserl?	40
& 4. Datos originarios y construcción teórica	46

CAPITULO 2.

PSICOLOGIA, METODO Y MENTE METODOLOGICA.

& 1. Presentación	51
& 2. Primeros escarceos	55
& 3. El método de la ciencia natural y el método de la psicología cognitiva	65
(a) El método de la ciencia natural: rasgos genéricos	65
(b) Modelos lingüísticos	68
& 4. El fisicalismo en psicología	79
& 5. Críticas al fisicalismo	89
(a) Planteamiento del problema	89
(b) Complicaciones ontológicas de toda posible superación del fisicalismo positivista en psicología	97
(c) Inconmensurabilidad nomológica entre sucesos	104
(d) El problema del dualismo causal	117
(e) Superveniencia	127
(f) Resumen	130
& 6. La situación experimental	134
& 7. Aplicaciones empíricas del método	142
(a) La representación de los significados	143
(b) Las post-imágenes que menguan	151
(c) "Aprendizaje" temporal visomotriz	154
& 8. Compromisos ontológicos	155

SEGUNDA PARTE.
LA SEMANTICA DE LAS ACTITUDES PROPOSICIONALES.

CAPITULO 3.
PERCEPCION, EXISTENCIA, REFERENCIA Y ERROR.

& 1. Prolegómenos	176
& 2. Yo veo, tú ves, él ve. Yo creo, tú crees, él cree. Materiales para una teoría intersubjetiva del conocimiento	179
& 3. El transfondo cognitivo y la causación sensorial	182
& 4. Solipsismo, independencia y covariación causal	190
& 5. Experiencias visuales, percepciones y creencias sobre existencias	195
& 6. Dos sentidos del verbo ver: <i>de re</i> y <i>de dicto</i>	205
& 7. Indexicalidad y percepción externa	208
& 8. Conocimiento por familiaridad y conocimiento por descripción . . .	211
& 9. La existencia y el problema del conocimiento	214
& 10. Imposibilidad de la duda absoluta en relación a la existencia de los objetos de percepción	217
& 11. Sobre percepción y existencia en la fenomenología	223
& 12. Los datos de los sentidos y el conocimiento del mundo externo . . .	229
& 13. Percepción y descripciones definidas	241
& 14. El rompecabezas de Russell	246
& 15. Nombres propios y descripciones definidas	248
& 16. Origen, cualidad y verdad de las descripciones	252
& 17. Nuevos ejemplos	259
& 18. La teoría de la referencia y sus dificultades	268
& 19. La solución del rompecabezas de Russell	270

CAPITULO 4.

LA TEORIA DE LAS ACTITUDES PROPOSICIONALES DE MARK RICHARD.

& 1. Hacia una teoría semántica global de las actitudes proposicionales . .	275
& 2. Decir con verdad lo que creen los demás	279
& 3. Los mediadores de las creencias	292
& 4. Aplicaciones de la teoría de los RAMs a algunos problemas semánticos	318
& 5. El aprendizaje y la retención de las creencias	324
Apéndice.	
& 6. La semántica y la ontología de las actitudes proposicionales	335

TERCERA PARTE.

EL PROBLEMA DEL SIGNIFICADO.

CAPITULO 5.

TEORIA CAUSAL DE LAS REPRESENTACIONES Y SISTEMAS REPRESENTACIONALES.

& 1. ¿Qué queremos de los estados mentales?	351
& 2. La teoría semántica naturalista de Dennis Stampe	353
& 3. Las tesis fundamentales	356
& 4. Sistemas representacionales del tipo I	359
& 5. Sistemas representacionales del tipo II	362
& 6. Teorías causales <i>versus</i> teorías ortodoxas	369
& 7. Sistemas representacionales del tipo III	380
& 8. Las representaciones lingüísticas	389

CAPITULO 6.
SOBRE CONTENIDOS Y REPRESENTACIONES MENTALES EN FODOR.

& 1. Hipótesis y límites de la investigación 398
& 2. Definición y división del problema 400
& 3. División del los conceptos 403
& 4. Los conceptos observacionales 404
& 5. Los conceptos teóricos 411

CUARTA PARTE.
EL PROBLEMA DE LA CONCIENCIA.

CAPITULO 7.
DOS TEORIAS DE LA CONCIENCIA: DENNETT Y SEARLE.

1. LA TEORIA DE DENNETT.

& 1. Psicólogos y filósofos 420
& 2. Tres clases de acceso a un contenido 426
& 3. Un mundo subpersonal para la conciencia 431
& 4. Heterofenomenología de la conciencia 447
& 5. La conciencia y el texto psicológico de Dennett 450
& 6. Rorty y el pragmatismo 457
& 7. Los hechos presionan de nuevo 460

2. INTENCIONALIDAD SIN CONCIENCIA:
BRENTANO, SEARLE Y LAS CIENCIAS COGNITIVAS.

& 1. Ciencias cognitivas e intencionalidad 471
& 2. La teoría de la intencionalidad de Searle 474

(a) Estados mentales y estados intencionales	474
(b) La estructura lógica de los estados intencionales conscientes . . .	476
(c) Intencionalidad intrínseca y conciencia	481
(d) La taxonomía de los estados mentales	483
& 3. Problemas de la teoría de la intencionalidad de Searle	486
& 4. El problema de Brentano	489
& 5. Subjetividad y neurofisiología	495
& 6. Resumen	498

CAPITULO 8.

HUSSERL Y LA CRITICA AL NATURALISMO EN PSICOLOGIA.

& 1. Fenomenología y filosofía objetivista	500
& 2. Actitudes recíprocas: de la fenomenología con la psicología científica, y viceversa	503
& 3. Mundo de la vida y ciencia	506
& 4. Necesidad de una psicología empírica	513
& 5. ¿Cómo hacer una psicología empírica?	518
& 6. El sujeto empírico y el sujeto trascendental	523
& 7. De la reducción trascendental a la conciencia absoluta	528
& 8. La crisis de la fenomenología trascendental y las ciencias europeas . .	542
& 9. La conciencia absoluta en el mundo	549
& 10. El dualismo epistemológico y el dualismo de las sustancias	553

BIBLIOGRAFIA	557
------------------------	-----

AGRADECIMIENTOS

Entre las personas e instituciones que han contribuido de una forma directa a la realización de esta Tesis Doctoral quiero mencionar especialmente y en primer lugar a su Director, Dr. D. Miguel García-Baró, por su generosidad, confianza y estímulo constante durante todos estos años. Asimismo doy las gracias a los Profesores D. Ramón Rodríguez García y D. Juan Miguel Palacios de la U.C.M. quienes hicieron posible una estancia en la Universidad de Wisconsin-Madison. También a esta misma universidad norteamericana por la concesión de una Beca (Research Assistanship) que me permitió trabajar en su Departamento de Filosofía durante el curso académico 1989-90. Para el Profesor Michael Byrd de la Universidad de Wisconsin-Madison toda mi gratitud y admiración por su apoyo y consejo desinteresado durante aquel año. Igualmente quiero dar las gracias a la Profesora Leora Weitzman por sus críticas y revisiones. Finalmente es imposible dejar de mencionar la ayuda inestimable que ha supuesto que pudiera disfrutar de una Beca Predoctoral Complutense entre los años 1990 y 1993.

PRIMERA PARTE.

**EN TORNO AL SER PSIQUICO:
SENTIDO COMUN, CIENCIA Y METACIENCIA.**

CAPITULO 1.

SENTIDO DE ESTAS INVESTIGACIONES.

& 1. El problema de Sócrates.

En el diálogo platónico *Fedón*, Sócrates cuenta la decepción que sintió en su juventud cuando se entregó al estudio de la teoría del filósofo Anaxágoras con el fin de penetrar en lo que hay de verdad en la tesis que afirma que la inteligencia (*Nous*) lo ordena todo y es la causa de todo (97 c). Para Sócrates, la consecuencia de esta afirmación sería que para cualquier objeto de estudio explicar su causa consistiría en explicar la necesidad de por qué una cosa es mejor que exista y que exista de determinada manera mejor que de otra cualquiera. Sin embargo, Sócrates se siente profundamente decepcionado cuando en el transcurso de la lectura del filósofo de Clazomene se da cuenta de que éste no atribuye a la "inteligencia ninguna causalidad en la ordenación de las cosas, sino que aduce como causas aires, éteres, aguas y otras muchas cosas absurdas" (98 b-c). Justo a continuación dice:

Me pareció que había algo muy parecido a como si uno afirmara que Sócrates hace todo lo que hace con inteligencia y, luego, al intentar exponer las causas de lo que hago, dijera que ahora estoy aquí sentado por esto, porque mi cuerpo está formado por huesos y tendones, y que mis huesos son sólidos y tienen articulaciones que los separan unos de otros....Así que al balancearse los huesos en

sus propias coyunturas, los nervios al relajarse y tensarse a su modo hacen que yo sea ahora capaz de flexionar mis piernas, y ésta es la razón por la que estoy yo aquí sentado con las piernas dobladas. Y a la vez, respecto de que yo dialogue con vosotros diría otras causas por el estilo, aduciendo sonidos, soplos, voces y otras mil cosas semejantes, descuidando nombrar las verdaderas causas: que una vez que a los atenienses les pareció mejor condenarme a muerte, por eso también a mí me pareció mejor estar aquí sentado, y más justo aguardar y soportar la pena que me imponen. Porque, ¡por el perro!, según yo opino, hace ya tiempo que estos tendones y estos huesos estarían en Mégara o Beocia, arrastrados por la esperanza de lo mejor, si no hubiera creído que es más justo y más noble soportar la pena que la ciudad ordena, cualquiera que sea, antes que huir y desertar. Pero llamar causas a las cosas de esa clase es demasiado absurdo. Si uno dijera que sin tener cosas semejantes, es decir, tendones y huesos y todo lo demás que tengo, no sería capaz de hacer lo que decido, diría cosas ciertas. Sin embargo, decir que hago lo que hago a causa de ellas, y eso al actuar con inteligencia, y no por la elección de lo mejor, sería un enorme y excesivo abuso de expresión. Pues eso es no ser capaz de distinguir que una cosa es lo que es la causa de las cosas y otra aquello sin lo cual la causa no podría nunca ser causa (Fedón 98c-99b).

Una mirada mínimamente atenta a la Historia de la Filosofía revela que el problema de entender y explicar cómo es posible que la inteligencia sea causa en seres que son materiales, o, como lo denominaré a partir de ahora siguiendo a Fodor (1987, 129), "el problema de la inteligencia encarnada", ha estado perpetuamente acuciando a las mentes filosóficas más sobresalientes de todos los siglos. Desde una de las primeras exposiciones más lúcidas y claras del problema—tal y como se encuentra en el texto de Platón que acabo de citar—, hasta terminar en las dos preguntas claves con las que ha de enfrentarse toda fundamentación de las ciencias cognitivas—el problema del significado y el problema de la conciencia—, el enigma de la inteligencia encarnada ha sido siempre uno de los misterios incommovibles de la realidad. Es él el que orienta la investigación que aquí se emprende. Todo el resto del capítulo tiene el objetivo prioritario de aclarar de una

forma más precisa cómo podría empezar a explorarse estos interrogantes y, en especial, cómo los ha encarado la filosofía de la mente contemporánea.

& 2. Sócrates y la caracterización del ser psíquico.

(a). Imprescindibilidad y profundidad de la psicología popular.

El comportamiento de Sócrates cuando conversaba con sus amigos el día anterior a su muerte, constituye un ejemplo inmortal y trágico de cómo cierta clase de representaciones mentales—las creencias, los deseos, los juicios, las valoraciones y otros estados representacionales semejantes—puede ser invocada para explicar las causas de la conducta. Sin embargo, es irremediable dejar ahora al arbitrio de cada cual que el significado de la expresión "representación mental" cobre un perfil más o menos exacto. Han sido tantas y tan variadas las generaciones de filósofos y de movimientos filosóficos que se han servido de este concepto, bien para denostarlo, bien para hacer un amplio uso positivo de él, que decidir ya en el comienzo mismo la utilización de un significado concreto de entre los múltiples posibles, sería de por sí una labor ingente. Además, seguir este camino podría llegar a ser sumamente dañino para la investigación posterior. No estaríamos seguros de hasta qué punto un determinado giro condicionaría excesivamente y, así, preestablecería en cierto sentido, qué es lo que se va a encontrar y cómo hay que conceptualizarlo. Para aclarar mínimamente cuál es el giro al que se quiere aludir en este momento inicial, sólo se hace referencia a las intuiciones más básicas que todos compartimos acerca de cómo explicamos cotidianamente alguno de nuestros comportamientos. Sencillamente, los explicamos como los explicaba Sócrates: recurrimos a las representaciones mentales y decimos que ellas están esencialmente involucradas en los procesos etiológicos que dan origen a nuestras

conductas. En este sentido, no hemos hecho grandes progresos a pesar de todo el tiempo que ha transcurrido desde la muerte de Sócrates.

La ahistoricidad que parece ser propia de esta clase de explicaciones causales de la conducta puede considerarse como un indicio de la utilidad y profundidad de la "teoría" que propone esta clase de explicaciones. Esta "teoría" recibe el nombre de "psicología popular" (Folk Psychology) en las discusiones actuales sobre filosofía de la mente. La psicología popular formaría parte de ese conjunto de estructuras conceptuales máximamente genéricas con las que todos contamos de antemano y que usamos para describirnos y describir el mundo en que vivimos y actuamos. Formaría parte de la "imagen manifiesta del Hombre-en-el-Mundo" en el sentido que dio Sellars (1963) a este término. Poner en claro cuáles son las entidades de las que se sirven las explicaciones de la psicología popular y cuáles son sus propiedades, supondría saber con qué dificultades esenciales tendría que enfrentarse una teoría psicológica que quisiera reivindicarlas como entidades teóricas dignas de una consideración científica madura y que no estuviera atendida meramente a lo que el sentido común dictamina. Uno de los propósitos de este capítulo es conseguir una cierta claridad sobre esta cuestión. Husserl, dentro de la tradición fenomenológica, y Fodor, Garfield y Pylyshyn, entre los filósofos de la mente contemporáneos, son algunos de los numerosos autores que han subrayado con más frecuencia la idea de que la inteligibilidad de la empresa psicológica científica y, así, el fundamento que le da sentido en cada uno de sus estadios, depende de la claridad que seamos capaces de hacer sobre este punto de partida. Por consiguiente, lo que está en juego aquí no es otra cosa que la aprehensión de la esencia genérica de los objetos que deben ser sometidos a los métodos de la investigación psicológica especializada.

Desde el punto de vista de la psicología popular, se puede establecer tentativamente que las creencias y deseos de Sócrates son representaciones

mentales, o, en terminología ya clásica, actitudes proposicionales. Por el contrario, los huesos y los tendones no lo son. Las actitudes proposicionales de Sócrates son, según él y según todos nosotros—al menos en principio—, "las verdaderas causas" de por qué está esperando tranquilamente la muerte en la ciudad de Atenas. En definitiva, entender la etiología de lo que él está haciendo, no es describir la constitución física de todas y cada una de las partes de su cuerpo. Tampoco bastaría con la formulación de las leyes de generalidad empírico-contingente que expresan las regularidades contrafactuales que se dan entre esas partes. Para entender su comportamiento, antes que con cualquier otra cosa, sería necesario contar con el conocimiento de sus actitudes proposicionales y asumir que al menos una parte de sus comportamientos está relacionada con sus representaciones mentales por medio de ciertos principios y leyes generales. La psicología popular está, por consiguiente, implícitamente comprometida con la tarea de proporcionar una explicación causal de la conducta y de las capacidades conductuales de los organismos que, quizá, sólo la psicología científica estaría en condiciones de formular explícita y sistemáticamente.

Las ventajas fundamentales con las que se nos presentan las explicaciones psicológicas de sentido común son su comprensibilidad y sus pretensiones predictivas. Aunque probablemente es verdad que la singularidad de Sócrates es irrepetible, sin embargo, ¿quién, cuando leía el texto de Platón, no ha pensado que si estuviera en posesión de los mismos deseos y creencias que Sócrates no estaría dispuesto a seguir su conducta? La clase de explicación de las causas del comportamiento que Sócrates considera adecuada tiene, desde este punto de vista, la propiedad de ser sumamente comprensible para todos nosotros. No es de extrañar. Estas explicaciones, como acaba de recordarse, las utilizamos constantemente en nuestra vida cotidiana. De hecho, cuando Platón escribió *Fedón*, seguramente contaba con que todos le entendiéramos desde un principio cuando su Sócrates negaba

que la disposición de los huesos y los tendones fuera la causa de su sobrecogedora espera. Sin embargo, su advertencia es también una señal inequívoca de que no todo el mundo en Grecia era un "realista intencional", es decir, que no todos admitían que la atribución de actitudes proposicionales a un sujeto fuera el medio más adecuado para explicar la etiología de su comportamiento.

Pero estas explicaciones psicológicas de sentido común no solamente son comprensibles, sino que tienen, como se ha mencionado, la pretensión de ser, en general, predictivas. Que lo sean, hasta qué punto lo sean, o cómo es posible que lo sean, son cuestiones mucho más difíciles de contestar y sobre las que se discutirá con detalle más adelante. Sin embargo, si la explicación que Sócrates nos da de su conducta (moral) no fuera comprensible y, hasta cierto punto, predictiva, ¿cómo sería posible que pensáramos de la forma en que lo hacemos cuando leemos aquel pasaje de *Fedón*? ¿No se convertiría todo él en una pura ininteligibilidad? De momento valga esta intuición mínima para dar por válida la plausibilidad de la afirmación de que las actitudes proposicionales y los comportamientos parecen estar relacionados por medio de ciertos principios y leyes generales. El conocimiento que de ellos tenemos--de una forma seguramente confusa--es el que permitiría que esta clase de explicaciones fuera inteligible y predictiva.

Sobre su carácter predictivo conviene, sin duda, decir algunas palabras más. Las explicaciones causales de la conducta que hacen uso de las representaciones mentales no solamente son las más apropiadas, según Sócrates, para entender la etiología de sus comportamientos, sino que también forman parte esencial del modo en que sus amigos son capaces de anticipar lo que Sócrates va a hacer. Fue suficiente que Sócrates les dijera que aceptaba la sentencia de los ciudadanos, para que ellos, después de dejarle solo durante unas horas, volvieran a la cárcel con la seguridad de

que seguiría allí esperando. Y, en efecto, Sócrates seguía allí esperando, a pesar de que, según cuenta Platón, no le hubiera resultado demasiado complicado escaparse de la cárcel. Las explicaciones psicológicas de sentido común parece que median de una forma sorprendentemente predictiva determinado grupo de interacciones que establecemos entre unos y otros. El problema de la coordinación de los comportamientos encuentra una solución sencilla en la psicología de sentido común (Fodor 1987, 3). La psicología de sentido común es tan estimada y usada por todos nosotros en nuestras interacciones de la vida cotidiana, porque una de sus funciones fundamentales es permitirnos hacer inferencias mediante las cuales predecimos las intenciones de las personas a partir de lo que ellas dicen, y porque parece que consiguen este objetivo con un grado de eficacia tan elevado como difícilmente superable por otras teorías. Por ejemplo, si Sócrates dijo que estaría en la cárcel algunas horas después, entonces es porque tenía la intención de estar en la cárcel unas horas después.

Las explicaciones psicológicas de sentido común también permiten hacer inferencias que permiten predecir muy fiablemente el comportamiento de las personas a partir de sus intenciones. Si Sócrates tiene la intención de estar en la cárcel unas horas después, entonces se comportará de una manera tal que, al final, resulte que unas horas después seguirá estando en la cárcel. La tesis que ya se está insinuando aquí es que Sócrates y sus amigos contaban con una teoría implícita--la psicología de sentido común--para hacer estas predicciones tan útiles. Si sus amigos no le hubieran encontrado en la cárcel, tal y como esperaban, su reacción habría sido más bien pensar que se lo habían llevado contra su voluntad a algún otro lugar, que atribuir aquel hecho a un fallo en el poder predictivo de la psicología de sentido común que implícitamente estaban usando.

Indudablemente, los susceptibles pensarán inmediatamente--no exentos de razón, por otra parte--que no hay que ir muy lejos para encontrar

ejemplos sencillos que refuten la consistencia y generalidad de esta clase de explicaciones. Para ellos, la conclusión más obvia que se desprende de esta especie de declaración de principio que se acaba de hacer acerca de las pretensiones y usos de las generalizaciones de la psicología popular, es que éstas son válidas sólo hasta un cierto límite. Sin embargo, como muy certeramente ha puesto de manifiesto Fodor, este límite está marcado, entre otras condiciones, por hechos que caen seguramente fuera del ámbito de aquéllos para los que formulamos con pretensiones de validez las generalizaciones de la psicología de sentido común. Ciertamente, nadie niega que haya numerosas excepciones a la clase de generalizaciones que formula la psicología popular. Sócrates dijo que estaría en la cárcel, aunque podría haber mentido, le podrían haber llevado a la fuerza a otro lugar, o haberle secuestrado. Sin embargo, esta cuestión acerca de la falibilidad y, por lo tanto, de la falsedad de las generalizaciones de la psicología popular, se analizará en breve con detenimiento y matizando los supuestos de los que depende. Pero antes una pequeña digresión necesaria.

(b). El vocabulario intencional.

Al hilo de la exposición, se ha hecho patente que la psicología popular, en la que normalmente todos confiamos de una manera más o menos ingenua para explicar nuestro comportamiento y el de nuestros congéneres, se caracteriza por servirse de un determinado vocabulario al que Pylyshyn denomina "vocabulario cognitivo" (1985, 2). Otra denominación que se utiliza normalmente para referirse a este tipo de léxico es la de "vocabulario intencional". Que a Sócrates le *parezca mejor* estar sentado en la cárcel, que *crea* que es justo soportar la pena, que *desea* llevar a cabo la condena que le han impuesto los ciudadanos, son expresiones en las que aparece típicamente este vocabulario intencional. Ha sido la constatación de

este hecho lo que ha llevado a considerar a muchos filósofos de la mente contemporáneos que el problema de la inteligencia encarnada es el nexo de unión de múltiples preocupaciones y tradiciones filosóficas aparentemente divergentes. Entre ellas cabría señalar fundamentalmente a la tradición fenomenológica, preocupada de una manera esencial por el problema de la intencionalidad, y a la tradición analítica, más preocupada, quizá, por la causalidad de los estados mentales. Es esta convergencia la que hace de él un problema tan sugerente y radical como extremadamente complicado.

La cuestión que se plantea ahora es sopesar hasta qué punto estamos autorizados para utilizar este vocabulario intencional a la hora de describir y explicar la conducta y, en especial, la conducta que llamamos racional. La utilización de este vocabulario estaría plenamente justificada si fuera imposible formular ciertas generalizaciones válidas y relevantes para las representaciones mentales, cuando éstas se invocan como causas del comportamiento, si se empleara cualquier otra clase de vocabulario. Desde luego, la psicología popular piensa que esto es lo que sucede en realidad.

Acertar en la elección del vocabulario del que ha de servirse una teoría para describir sus objetos y expresar sus leyes no es, en modo alguno, una cuestión superficial. El vocabulario en el que se enuncian las leyes de una ciencia es esencialmente el mismo vocabulario que se utiliza para describir los fenómenos que han de ser explicados y que caen dentro de su dominio. Sería absurdo intentar una explicación de hechos geológicos, como son las fallas y los terremotos, enunciando leyes geológicas que se expresaran, por ejemplo, en términos bioquímicos o neurofisiológicos. Por consiguiente, una mala elección del vocabulario de una ciencia puede tener consecuencias nefastas en el desarrollo posterior de esa ciencia. La razón básica para hacer esta afirmación es que la elección del vocabulario de una ciencia está en conexión esencial con su idea, esto es, con la caracterización previa de las propiedades de los objetos que pretende estudiar. Por este

motivo, quien negase que el vocabulario intencional es imprescindible para describir la esencia genérica de los hechos psicológicos y, de acuerdo con la psicología popular, que él es el vocabulario adecuado para explicar las causas del comportamiento, negaría en bloque la idea de una psicología que cuantificara sobre objetos con propiedades intencionales, al tiempo que rechazaría cualquier orientación de la investigación psicológica que quisiera seguir los derroteros indicados por la psicología popular. Para este hipotético individuo, afirmar que la psicología científica debe apoyarse, al menos en su inicio, en los datos aportados por la psicología popular, equivaldría a decir el mismo absurdo que afirmar que la biología es la ciencia de los animales con corazón y de las piedras. En ambos casos se estaría cometiendo la falacia conocida con el nombre de "metábasis eis allo genos".

Con los conocimientos actuales sería difícil negar que una base descriptiva de los hechos que son el objeto de estudio de la ciencia biológica en la que aparecieran con el mismo derecho términos como "animal con corazón" y "cuarzo" y sobre cuyos referentes se quisieran formular generalizaciones que los incluyera por igual, haría de esa hipotética ciencia biológica un cúmulo de supersticiones y sinsentidos. En este caso se habrían confundido dos especies de dos géneros distintos pero cuyos individuos se considerarían ejemplares adecuados del dominio de una misma ciencia. Ahora es fácil saber de cuál de las dos especies hay que desprenderse.

Paralelamente, diría aquel supuesto interlocutor, si hay una ciencia cuyo tema es la investigación de las causas de la conducta, entonces, para esa ciencia, debe quedar descartado el uso del vocabulario intencional, así como quedaba descartado el uso del término "cuarzo" del vocabulario de la biología. Para explicar las causas, la ciencia sólo precisa de términos como "huesos" y "tendones", y no de términos como "creencia" y "deseo".

Lo que habría que probar contra este hipotético oponente es que para elaborar una taxonomía adecuada de los comportamientos, objeto por antonomasia de la psicología popular, se necesita esa clase de léxico intencional, y no tanto el léxico físico.

El argumento que se suele ofrecer en este punto es siempre el mismo: las acciones clasifican transversalmente los movimientos. (Véase, Pylyshyn (1985) y Fodor (1987)). Por ejemplo, la acción que consiste en ir a la tienda de ultramarinos para comprar azúcar—y que así se describe—, se puede realizar físicamente de múltiples maneras. Se puede ir a la tienda andando, en coche, en autobús o en bicicleta. La descripción de todos los variadísimos *tipos de movimientos* en que consiste andar, ir en coche o en bicicleta, no puede constituir propiamente la descripción adecuada del *tipo de acción* que el sujeto realiza. El tipo de acción que un sujeto realiza depende esencialmente de las intenciones que tenga. En este caso, la acción está determinada por la intención que se describe como "ir a la tienda de ultramarinos para comprar azúcar". Así, pues, en oposición a quien argumenta en contra de la legitimidad con la que se utiliza en psicología el vocabulario cognitivo en favor del vocabulario físico, anatómico o neurofisiológico, la moraleja de este ejemplo es que sólo bajo la descripción "ir a la tienda de ultramarinos para comprar azúcar" se puede hablar del comportamiento de ese sujeto, y no bajo la descripción "subir los peldaños de un autobús" o "mover las piernas circularmente para dar pedales". Por consiguiente, no cabe elaborar una taxonomía del comportamiento sin el léxico intencional. Si hay generalizaciones válidas que expliquen las causas de la conducta, tendrán que formularse haciendo uso de él, dado que la descripción de los comportamientos que caen bajo aquellas generalizaciones debe hacerse, al menos en principio, empleando dicho vocabulario. Con todo, estas afirmaciones no excluyen de antemano, claro está, que realizar aquella acción no sea realizar al menos uno de esos tipos de movimientos,

aunque la acción no se pueda identificar propiamente con ninguno de ellos, ni con su conjunción.

Desde esta perspectiva, no es otra la cuestión con la que se enfrentaba Sócrates que la de establecer cuál es el vocabulario adecuado para describir la circunstancia en la que se encontraba. Sócrates diría hoy que los vocabularios de la física, de la química, de la neurofisiología son, desde luego, adecuados tanto para describir un determinado suceso físico, químico o neurofisiológico, como para describir la cadena de sucesos antecedentes que son causalmente responsables de ellos. Sin embargo, también diría que las leyes empírico-contingentes que podamos enunciar sobre estos sucesos son solamente relevantes para ellos bajo esas descripciones: las piernas y los tendones están doblados porque desde el punto de vista químico o neurofisiológico pasa esto, esto y esto. Sin embargo, una respuesta de este último estilo dejaría bastante insatisfecho a Sócrates. Esta respuesta podría constituir una explicación a la pregunta ¿por qué este cuerpo está en este lugar en esta posición? Pero no una explicación a la pregunta ¿por qué Sócrates está sentado en la cárcel? La contestación de la psicología popular a esta segunda cuestión ya la sabemos.

Con todo, y aquí anida un problema esencial que aparecerá sucesiva e insistentemente, lo que tiene que decir la neurofisiología acerca de la etiología del suceso "este cuerpo--el de Sócrates--en esta posición", seguramente debería formar parte de la explicación global de las condiciones que hacen posible que un organismo lleve a cabo las acciones que menciona la explicación psicológica de sentido común (Pylyshyn 1985, 4). Ya lo dijo Sócrates.

Así, pues, la psicología popular, tal y como se ha explicado, considera que el vocabulario cognitivo no es simplemente un artilugio conceptual pasajero que puede servir de ayuda como método heurístico para dar cuenta de las causas de ciertos comportamientos. No forma parte de

las pretensiones explicativas de la psicología popular el verse a sí misma bajo la forma de la provisionalidad hasta que el desarrollo de otras ciencias, como, por ejemplo, la neurofisiología o el conductismo, haga que éstas alcancen sus metas teóricas. La psicología popular cree firmemente que la principal razón de su existencia en el reino de las teorías es que solamente ella puede dar cuenta de ciertas regularidades que son relevantes para una determinada esfera de objetos, es decir, que las generalizaciones que formula para esos objetos son imprescindibles. Ninguna otra teoría puede hacer esto sin cometer una "metábasis eis allo genos".

(c) *Las cláusulas ceteris paribus.*

Las generalizaciones que la psicología popular es capaz de enunciar haciendo uso del vocabulario intencional, sólo empiezan a adquirir el carácter de una explicación legítima del comportamiento cuando se expresan mediante los característicos enunciados condicionales contrafactuales. Si el comportamiento de Sócrates es realmente racional, sistemático y regular, entonces un cambio en sus actitudes proposicionales tendría, o podría tener, como consecuencia un cambio correspondiente en su conducta. De haber tenido otras creencias y otros deseos, quizá hubiera puesto pies en polvorosa camino de Mégara. Una marca esencial de las explicaciones causales de la psicología popular, en oposición a las meras descripciones, es que los enunciados que expresan las primeras siguen el esquema general: "Si *X* no hubiera creído *Z* y deseado *Y*, y..., entonces no hubiera hecho *C*". (Los puntos suspensivos son, ciertamente, imprescindibles. El antecedente de la cláusula condicional debe ser completado con nuevas condiciones cuya constatación y enumeración completa puede ser un asunto tan delicado como complejo. Todos sabemos el abismo que hay, muchas veces imposible de superar, entre la realidad y el deseo.) Los puntos suspensivos están en

lugar de las cláusulas *ceteris paribus*. El rasgo específico de las generalizaciones en las que aparecen estas cláusulas—como es manifiesto en las generalizaciones implícitas de la psicología del sentido común—, es que quedarían libres de su lastre si se pudieran enunciar todas las condiciones de las que depende su validez. Una ciencia en la que se pudiera conseguir hacer esto sería, según la terminología de Donald Davidson, una ciencia "perfectible" (Fodor 1987, 5). Según este criterio de Davidson, las ciencias reales serían perfectibles, aunque no las ciencias intencionales, como la psicología. Pero, ¿es cierto que no cabe incluir a la psicología entre las ciencias perfectibles? Si la psicología popular no es una ciencia perfectible porque nunca podría deshacerse de las cláusulas *ceteris paribus*, tampoco lo son ninguna de las ciencias reales que llamamos "ciencias especiales". Entre éstas sólo la física lo sería. Esta es la tesis de Fodor.

Para ilustrarla, este filósofo pone el ejemplo de las generalizaciones de la ciencia geológica. Una generalización sumamente modesta de la geología podría ser ésta: "un río que tiene meandros erosiona la orilla externa". El problema de esta clase de generalizaciones es que sin la cláusula *ceteris paribus* parece ser claramente falsa, pero con ella no tiene valor explicativo alguno. Sin ella es falsa porque, por ejemplo, el tiempo puede cambiar y ocasionar que el río se seque o se congele; los ingenieros pueden construir una presa y contener el agua del río; se puede acabar el mundo, etcétera. Cualquiera de estos sucesos falsearía la generalización porque evitaría que el río erosionara orilla alguna. Para evitar este inconveniente, se puede intentar reformular el enunciado que expresa la generalización de esta otra manera: (i) "Si el resto de las condiciones permanece igual, un río que tiene meandros erosiona la orilla externa". Pero este enunciado, se podría pensar, no explica absolutamente nada que sea geológicamente relevante porque equivaldría a decir tanto como que (ii) "un río que tiene meandros erosiona la orilla externa a menos que no lo

haga". Este enunciado tiene el problema de que es explicativo de un modo muy poco interesante para la ciencia geológica: nada que suceda podría hacerlo falso (Fodor 1987, 5).

Sin embargo, no parece que (i) y (ii) signifiquen exactamente lo mismo, según considera Fodor estos enunciados. Enunciados como (i) son informativos aunque abiertos a la posibilidad de excepciones. En ellos se quiere dar a entender que "*P* es el caso en todo mundo posible para el que sean válidas las idealizaciones propias de la geología". Los enunciados de las ciencias especiales, ciertamente, pertenecen a esta clase. Por el contrario, enunciados como (ii) simplemente significan algo así como que "*P* es el caso en todo mundo posible a menos que *no P*". A nadie le puede interesar demasiado formular enunciados de este estilo para ninguna región posible de objetos naturales. Es manifiesto, por consiguiente, que lo que se dice en (i) es mucho más que lo que se dice en (ii) (Fodor 1987, 5). La conclusión es que si la objeción contra la validez de las generalizaciones de la psicología popular está basada en que todas ellas se expresan mediante enunciados contrafactuales que contienen cláusulas *ceteris paribus*, lo mismo sucede con el resto de las generalizaciones de las ciencias especiales. Todas ellas tienen excepciones si se suprimen estas cláusulas. Todas menos, quizás, la física idealmente completada.

Antes de terminar con estas aclaraciones imprescindibles, conviene mencionar explícitamente que una característica de los sucesos que hacen falsas las generalizaciones de las ciencias especiales, es que estos sucesos no se pueden describir haciendo uso del mismo vocabulario que se emplea cuando se describen los sucesos para los que sirven contingentemente aquellas generalizaciones. Por ejemplo, en enunciados como "un río que tiene meandros erosiona su orilla externa a menos que se contruya una presa" o "Sócrates permanecerá en la cárcel a menos que le secuestren o que sus centinelas le lleven a la fuerza a otro lugar", ni "construir una

presa" ni "secuestrar" o "llevar a la fuerza a otro lugar", son términos del vocabulario de la geología o de la psicología, respectivamente.

Los sucesos que constituyen excepciones a las generalizaciones de las ciencias especiales se describen, usualmente, *desde otra ciencia*. El caso más frecuente es aquel en el que se desciende uno o más niveles en la jerarquía del sistema de las ciencias para expresar estas excepciones. De la geología descendemos a la ingeniería, de la psicología popular a la criminología. Las cláusulas *ceteris paribus* son ineliminables de las ciencias especiales, aunque sean detallables y explicables en el vocabulario de alguna otra ciencia más básica (Fodor 1987, 6).

(d) Intencionalidad y causalidad.

Sin embargo, estas primeras aproximaciones al planteamiento general del problema de la inteligencia encarnada no son todavía suficientes. Si Sócrates hubiera deseado marcharse de la ciudad en vez de soportar la pena, y hubiera creído que así podía hacerlo, su comportamiento, permaneciendo el resto de las condiciones iguales, hubiera sido, muy probablemente, distinto. Esto no es nada más que una forma técnica de decir que cuando un sujeto varía sus creencias y deseos varía, seguramente, su comportamiento. Esta indicación apunta hacia el hecho de que las actitudes proposicionales deben de portar algún rasgo que las diferencie entre sí y que, en algún sentido, sea responsable de los efectos que ellas tienen en la conducta del sujeto a quien se adscriben. La literatura filosófica denomina a este rasgo diferenciador de múltiples maneras. Las más comunes son las de "contenido semántico", "contenido intencional", "objeto" o "proposición". De cualquier modo, todas estas caracterizaciones, que en algunos casos pueden dar lugar a ciertos equívocos, sólo vienen a insistir en la idea de que las actitudes proposicionales son estados intencionales.

Otra forma de expresar esta misma idea es diciendo que las actitudes proposicionales tienen significado o, incluso, que ellas son, propiamente, las portadoras intrínsecas de significado.

El problema de la inteligencia encarnada adquiere ahora un nuevo matiz antes desconocido. De acuerdo con esta última consideración, el problema de la inteligencia encarnada se concreta en buscar las condiciones de posibilidad para que podamos decir justificadamente que las causas de verdad de la conducta ejemplar de Sócrates son sus actitudes proposicionales en virtud de poseer determinados contenidos semánticos o intencionales. Por contenido semántico o intencional de una actitud proposicional hay que entender en un principio aquello a lo que se refieren las oraciones que en los lenguajes naturales funcionan como complementos directos de verbos que expresan actitudes proposicionales, es decir, aquello a lo que se refieren los complementos gramaticales de oraciones del tipo, "Sócrates cree que ...", "Sócrates desea que ...", "Sócrates espera que ...", etcétera. Como mínimo, es la conjunción de estas dos propiedades fundamentales de las actitudes proposicionales--su capacidad causal y sus contenidos intencionales--la que explicaría el hecho de que Sócrates esté sentado en una cárcel de Atenas y no camino de Mégara. Sócrates podría haber creído que lo mejor era cumplir con la sentencia dictada por los ciudadanos de Atenas, pero no desear cumplirla. Si ésta hubiera sido la situación psíquica en la que se hubiera encontrado, la interacción causal de sus creencias y deseos habría dado lugar, *ceteris paribus*, a una conducta muy distinta de la que nos narra Platón.

Precisamente, es un lugar común que la individuación de las actitudes proposicionales se lleva a cabo señalando quién es su portador, de qué clase de actitud se trata en cada caso y cuál es su contenido intencional. En un principio, las combinaciones posibles entre los sistemas intencionales (los portadores de actitudes proposicionales, los "sujetos"), sus

actitudes proposicionales y los objetos de éstas, son múltiples. Así, no parece absurdo pensar que distintos sistemas intencionales puedan estar en el mismo estado intencional y estar dirigidos hacia el mismo objeto. Expresiones en el lenguaje cotidiano del tipo "José Manuel y Alicia creen que la película empieza a las 7'30", en las que se adscribe la misma creencia a dos sujetos, son moneda corriente. Por otra parte, se puede tener la misma actitud proposicional respecto de muy variados objetos. Se puede desear comer una manzana, ir al cine, quedarse en casa o que escampe. Finalmente, se pueden tener actitudes proposicionales muy diversas hacia el mismo objeto. Se puede creer que la película empieza a las 7'30, desear que la película empiece a las 7'30, esperar que la película empiece a las 7'30 y todas las combinaciones que sean posibles con estos elementos.

Según Fodor, sin duda uno de los autores que más ha tratado de perfilar las propiedades del ser psíquico que defiende la psicología popular y a quien hemos seguido de cerca en estas primeras consideraciones, esta clase de reflexiones da pie a pensar que los estados mentales que son muestras de tipos de actitudes proposicionales deben llegar a poder formar *entramados de causas y efectos* si es que es verdad que las actitudes proposicionales deben ser tenidas especialmente en cuenta a la hora de explicar las causas de las conductas intencionales de los sujetos. Estos entramados de causas y efectos entre los diversos estados mentales que forman nuestra vida psíquica se llaman normalmente *procesos mentales*. Lo propio es que pensemos naturalmente en dichos procesos mentales como las causas (eficientes) de muchos de nuestros comportamientos. Las actitudes proposicionales serían los eslabones imprescindibles de estas cadenas causales que, finalmente, originarían el comportamiento manifiesto. Pero formarían parte de ellas de una manera muy peculiar, al menos desde el punto de vista de la psicología popular. Las actitudes proposicionales no sólo conservan, claro está, las relaciones semánticas entre sus contenidos

intencionales, sino que, además, se las apañan de alguna manera para interactuar causalmente entre sí.

Es en los silogismos prácticos donde se pone de relieve de una manera paradigmática la forma de pensar más común en la que las actitudes proposicionales respetan relaciones semánticas características y constituyen, a la vez, entramados de secuencias causales. El ejemplo hipotético de Sócrates que se ofrecía hace tres párrafos podría servir muy bien como ilustración de la manera en que los silogismos prácticos expresan el isomorfismo entre las relaciones semánticas y las relaciones causales de las actitudes proposicionales¹.

Por informales que parezcan en principio estas precisiones, sin embargo, lo que ellas ponen de manifiesto es que las explicaciones de la psicología popular apelan esencialmente a entidades a las que atribuyen simultáneamente poderes causales y contenidos semánticos. Estas son las propiedades esenciales de las entidades que caen bajo las generalizaciones de la psicología popular y de las que estaría hablando implícitamente Sócrates. Son también la clase de entidades a las que debería atender toda psicología científica que considere que la psicología popular es un punto de partida legítimo para sus investigaciones especializadas.

Como se puede suponer fácilmente, las preguntas a que darían origen estas características de las entidades de las que habla implícitamente la psicología popular son metafísicamente muy interesantes. El problema de la inteligencia encarnada se concreta ahora en saber cómo podría estar

¹ El problema del isomorfismo que aquí se plantea se discutirá y desarrollará más pormenorizadamente en el § 8 del capítulo segundo cuando se cuente con más elementos de juicio.

construida una mente o un "alma" para que los estados mentales pudieran estar relacionados tanto semántica como causalmente (Fodor 1987, 14)².

(e) La "promiscuidad" ontológica del ser psíquico.

El origen del rompecabezas teórico que Sócrates plantea es que, por una parte, la aprehensión precientífica del ser psíquico obliga a poner en marcha, como su motivación inteligible más originaria, una teoría psicológica acerca de los procesos mentales que explique el isomorfismo entre las relaciones semánticas y las relaciones causales que la psicología de sentido común atribuye a las actitudes proposicionales como causas del comportamiento. Por otra parte, la escisión dualista entre naturaleza en sí y mundo subjetivo que introdujo el avance de la metodología matemática, hace evidente que la psicología, como ciencia del ente psíquico, fuera automática y sistemáticamente descartada del reino de las ciencias "exactas" de la naturaleza. El desarrollo de la historia de la psicología desde la escisión dualista llevada a cabo en el pensamiento de Descartes, puede contemplarse como el intento denodado de la psicología por encontrar un método y una definición de sus problemas que le devuelva el derecho de pertenecer, sin asomo de duda, a las ciencias naturales especiales. Ahora, dice la psicología cognitiva, parece que se ha encontrado la idea que se necesitaba para explicar el isomorfismo entre relaciones semánticas y causales. Esta es, precisamente, la misma idea que ha de devolver a la psicología la pretensión, ahora ya sancionada, de ser una ciencia más entre

² Indudablemente, se podría introducir aquí alguna clase de consideración histórica acerca de la multiplicidad y relativo refinamiento progresivo de las "psicologías populares" que ha habido desde Homero a... ¿Heidegger? Esta tarea, sin embargo, está fuera del alcance de esta investigación. Quien esté insatisfecho con lo que se ha dicho hasta ahora puede considerar lo siguiente. Aquí no se ofrece la psicología popular *tout court*, sino sólo algunas de sus propiedades esenciales, aquellas cuyo rechazo supondría "la mayor catástrofe intelectual en la historia de nuestra especie" (Fodor 1987, xii)

las ciencias naturales. Esta idea feliz es el computacionalismo de cuño funcionalista. Por consiguiente, nada de psicología psicofísica, en último término genética, ni de psicología descriptiva, sino psicología cognitiva: ni genética plenamente, ni tampoco exclusivamente descriptiva. (Habrá ocasión de hablar pormenorizadamente de este problema en el siguiente capítulo.)

La importancia que tiene esta idea se hace claramente patente si se considera que el problema de Sócrates se complica mucho más cuando se tiene en cuenta que la psicología popular parece obligar a postular una ontología promiscua para el ser psíquico—el ser psíquico tiene propiedades semánticas y también causales—, pero que el mundo constituido y pensado con categorías científico-naturales no admite esta promiscuidad debido al dualismo inevitable a que dieron lugar la matematización de la naturaleza y la posterior eliminación sistemática de lo propiamente psíquico del reino científico. La imagen científica del mundo dice que si para hacer psicología debemos contar con un ser promiscuo, entonces no tendremos psicología. Tendremos neurofisiología o, finalmente, física. En cualquier caso, intentaremos reducir la psicología a otra ciencia más básica—posiblemente, en último término, a la física. Es decir, buscaremos teorías causales del comportamiento en las que no haya rastro de ninguna entidad con propiedades intencionales. Lo que sea preciso, antes que sucumbir, por así decir, a la "amoralidad" metafísica a propósito del ser psíquico. No seremos realistas intencionales.

Una razón, ya anticipada, para no ser realistas intencionales, se puede encontrar con cierta facilidad si se fuerza un poco el problema que plantea Sócrates. Aunque la psicología popular defiende que las actitudes proposicionales son esenciales para explicar la etiología de nuestros comportamientos, con todo, tiene que haber alguna clase de relación entre las causas y procesos mentales y "aquello sin lo cual la causa nunca podría ser causa". Si no fuera éste el caso, entonces la causa verdadera se quedaría

en mera posibilidad de actuar sobre "el mundo" sin poder tener jamás sobre él efecto real alguno. Lo que resulta un tanto embarazoso y hace del problema de la inteligencia encarnada un auténtico rompecabezas metafísico, es pensar que, además, las condiciones que permiten a la causa "inteligente" ser causa, son ellas mismas, a su vez, causa. Según la visión científico-objetiva, existen leyes empírico-contingentes que explicarían las disposiciones de los huesos y tendones a lo largo del espacio y el tiempo en función de sus condiciones espacio-temporales previas. Sin embargo, la esperanza ilimitada en la capacidad predictiva de la ciencia natural hace que concibamos la posibilidad bien fundamentada de llegar a explicar de una manera "mecánica" que Sócrates esté donde está en el preciso momento en que lo está sin apelar a las actitudes proposicionales. Todo lo que necesitaríamos para explicar su comportamiento sería la neurofisiología o, finalmente, la física--si es que uno sigue conservando una fe absoluta en el programa de la unidad de la ciencia, cosa extraña en estos tiempos. Si, a pesar de todo, Sócrates está empeñado en sostener que una explicación mecánica de su comportamiento es absolutamente inadecuada, entonces parece que estamos ante dos series causales paralelas que desembocan en la producción del mismo efecto. Pero ¿no supone esto una sobredeterminación causal del efecto? El problema es, por consiguiente, doble. Primeramente hay que preguntar cómo es en general posible que un sistema físico, como lo es el cuerpo de Sócrates, actúe "movido causalmente por meras representaciones mentales". La segunda cuestión relevante es ésta: ¿no es superflua cualquier teoría sobre la causación de las representaciones mentales cuando podemos explicar el mismo suceso sin necesidad de postular estas entidades "mentales"? Esta es una opción abierta que, quizás, no cabe descartar *a priori*. Entre los filósofos de la mente actuales más destacados han sido Paul y Patricia Churchland algunos de sus defensores más

entusiastas. Para ellos, en la "imagen científica del Hombre-en-el-Mundo" no hay lugar para categorías intencionales.

Es importante dejar anunciado ya aquí que el reto del punto de vista del programa de Fodor en filosofía de la psicología, del que nos ocuparemos en extenso en el próximo capítulo, es aceptar como punto de partida de la investigación psicológica científica la ontología que reclama para sí la psicología popular y, sin embargo, mantener también que lo que hay verdaderamente, el ser verdadero, es el que nos obligan a reconocer los compromisos ontológicos de las ciencias maduras y desarrolladas. Para el caso del ser psíquico, éstos son los compromisos ontológicos de la psicología cognitiva. La psicología y, en concreto, la psicología cognitiva, es la ciencia que por derecho propio se erigiría actualmente en árbitro último de la ontología de lo mental.

Creo que éste es un punto de importancia ilimitada en lo que atañe a la metodología y a los criterios que debemos utilizar para establecer las propiedades ontológicas reales de lo mental. Estas no podrían inscribirse en ninguna lista definitiva de lo que hay, antes de aplicar los baremos de evaluación epistemológica a la psicología que se presente paradigmáticamente como genuinamente científica, y antes de que reconozcamos las entidades sobre las que cuantifica dicha teoría. El único *a priori* ontológico que estaría en condiciones de ser puesto en la cabecera de aquella lista antes de cualquier discusión detallada, sería el que la psicología popular provea. Este *a priori* circunscribiría, bien provisionalmente, bien definitivamente—y decantarse por uno de los miembros de esta disyunción es una decisión crucial—, la región o dominio de objetos a los que debe aplicar sus diligencias la investigación especializada. En las primeras disquisiciones del capítulo se aludió al sentido que tenía la empresa de reconocer las propiedades del ser psíquico que postula la psicología popular. Este sentido no era otro que el de intentar una primera aprehensión de la esencia

genérica de este ser. Esta intención, sin embargo, se tiene que poder compaginar con el hecho histórico manifiesto de que no todas las ciencias consiguieron una definición de sus objetos y un planteamiento de sus tareas que fueran adecuados desde un inicio. Así, pues, el carácter provisional o *sub specie aeternitatis* que estemos dispuestos a conceder a la clase de evidencia con la que la psicología popular nos ofrece las propiedades del ser psíquico, es una elección que depende de las concesiones que se esté dispuesto a hacer a la validez de argumentos de este tipo:

Un dominio es una esfera de fenómenos...sobre los que "es posible desarrollar teorías basadas en un vocabulario especial o en un conjunto razonablemente uniforme de principios" distintos e independientes de aquellos que son apropiados para otros dominios. Los dominios, por tanto, no se individualizan directamente por referencia a una taxonomía arbitraria de las cosas que están contenidas en ellos..., sino más bien por referencia a los vocabularios y estrategias explicativas que son apropiados a su investigación. El inventario de las cosas en un dominio está determinado, a su vez, por los límites de una aplicación con éxito de estos vocabularios y estrategias. La individuación de dominios, por consiguiente, procede por principios pragmáticos más bien que por principios ontológicos. (Garfield 1988, 11-12).

Estas consideraciones ponen ante la vista del pensamiento filosófico que las propiedades del ser psíquico no pueden, quizá, ser establecidas *a priori* más allá de los compromisos ontológicos que se contraen cuando se aceptan las explicaciones psicológicas de sentido común. Lo interesante de la tesis de Fodor es que habría una ciencia que reclamaría para sí las propiedades del ser psíquico que la psicología de sentido común postula, pero que para dar cuenta con detalle y coherentemente de ellas, avanza en construcciones teóricas que de un modo cada vez más claro otorgan a sus objetos nuevas propiedades que jamás podrían haber sido postuladas ni sospechadas antes por la psicología popular. Precisamente, es esto último lo

que hace divertido dedicarse a la psicología científica, como dice Fodor con buen humor. Por lo demás, en este respecto al menos, no hay ninguna diferencia en relación al desarrollo, por ejemplo, de la ciencia física. Tampoco las propiedades de sus objetos se parecen en nada a las propiedades de los objetos físicos espacio-temporales de tamaño medio tal y como se nos presentan y se nos dan previamente en la "física popular".

La imagen manifiesta contiene los recursos conceptuales para plantear problemas para los que carece de recursos con que solucionarlos. (Garfield 1988, 109).

Se podría admitir, pues, que la "ampliación" y "transformación" de los objetos de la psicología popular por parte de la ciencia psicológica madura han de remitir siempre a las propiedades que estos objetos tienen en la psicología popular. La psicología popular sería el fundamento inteligible de la psicología científica. Sin embargo, en la reflexión filosófica posterior sobre las propiedades así "ampliadas" y "transformadas" será cada vez más difícil entender cómo es posible que haya cosas tales como las que postula la psicología popular. A la luz de estas reflexiones, creo que es de utilidad distinguir entre el fundamento de la inteligibilidad de una ciencia y su fundamento ontológico. La psicología popular, tal y como ha ido apareciendo a lo largo de estas discusiones, sería el fundamento inteligible de la psicología científica. Hablar del fundamento de la inteligibilidad de una ciencia no es otra cosa que remitirse a todas aquellas descripciones que parecen dar sentido a la tarea de una ciencia y que son accesibles en todo momento, y sin mayores explicaciones, tanto para los cultivadores de la ciencia, como para los legos. Es aquí, en todo caso, donde se podría hablar de un auténtico *a priori* inteligible de carácter ahistórico. El ejemplo de Sócrates lo pondría así de manifiesto. Sin embargo, las descripciones que forman parte del fundamento inteligible de una ciencia imponen ciertas

propiedades a las entidades sobre las que recaen sus generalizaciones y predicciones. La cuestión espinosa es que no se puede contar desde el comienzo con que estas propiedades sean definitorias del género de objetos que definen el dominio de una ciencia. Es posible que no haya ningún *a priori* ontológico de carácter ahistórico para ninguna región de objetos de la ciencia natural. Si en el límite ideal del acabamiento perfecto de las ciencias especiales los compromisos ontológicos que estemos obligados a contraer cuando las aceptamos coincidieran en un solapamiento perfecto con aquellos que contraemos cuando explicitamos sus fundamentos inteligibles, ésta sería una coincidencia imprevisible. Lo normal, por otra parte, es que no se dé siquiera este solapamiento. El desarrollo histórico de las ciencias especiales nos impele a reconocer este hecho. El caso de la psicología cognitiva y de sus compromisos ontológicos, desde luego, no sería una excepción a este estado de cosas.

& 3. Husserl y la caracterización del ser psíquico.

(a) Intencionalidad e "imperar" psíquico.

¿Puede ser cuestionado el punto de partida mismo de la investigación psicológica? Los dos datos primarios que se nos daban como punto de partida de la investigación psicológica--la intencionalidad y la causalidad de las actitudes proposicionales--¿son realmente primarios? Estas dos preguntas no se refieren ahora mismo a la posibilidad de erigir una "psicología neurofisiológica" que no dé cabida a entidades intencionales, sino sólo causales. Por el contrario, aluden a la posibilidad de construir una psicología exclusivamente intencional pero no causal, sin que esta psicología deje de ser una psicología científica. Es más, con la pretensión de que ella sea la psicología realmente científica: una psicología absolutamente carente

de ingenuidades y supuestos, que estaría alentada, para alcanzar esta meta, por un sentido de cientificidad radicalmente distinto del de la cientificidad natural-objetiva. Esta tercera opción está representada por las tesis de Husserl. Antes de continuar, sin embargo, es importante dejar constancia de que la posibilidad de plantearse estos interrogantes da cuenta del hecho de que la descripción de la imagen manifiesta del mundo no es tan estática como podría pensarse. (Una observación semejante hace Garfield (1988).)

Husserl tuvo el mérito enorme de enunciar con singular fuerza y nitidez por qué una psicología construida con los cánones de la cientificidad instaurados en la modernidad nunca podría haber servido de ayuda especial a la verdadera psicología, que, a la larga y mediante la aclaración constante de sus supuestos, debe identificarse con la filosofía transcendental. En este sentido, el dualismo psicofísico fue desastroso tanto para la filosofía como para la auténtica psicología. Husserl denunció este hecho pero no contestó específicamente cómo podría la psicología superar este dualismo congénito. Mejor dicho, sí que declaró cómo superarlo, a saber: eliminando el problema de la "inteligencia encarnada", en el sentido en el que hoy se entiende éste, mediante el planteamiento de la tarea de una investigación rigurosa de la subjetividad universal constituyente. Sin poder entrar directamente ahora en el género de investigaciones que debería abrazar este proyecto de Husserl, del que me ocuparé por extenso en el último capítulo, creo que el camino que quiso seguir Husserl estaba orientado por este otro lema: si para hacer psicología debemos contar con un ser promiscuo, entonces haremos psicología sin causalidad. Como se impone de suyo, la primera tarea será la de mostrar que no es correcto hablar de causalidad cuando se hace referencia a las propiedades ontológicas del ser psíquico tal y como se da éste en la psicología popular, en el

"mundo de la vida" (*Lebenswelt*), como dice Husserl³. Para hacer posible esta eliminación, el primer paso será intentar caracterizar el ser psíquico--las almas, las mentes--tal y como se nos da previamente en el mundo de la vida. Esta caracterización debe ser llevada a cabo, según Husserl, "de una manera realmente desprejuiciada,... de un modo totalmente ametafísico" si es que se quiere "capturar el último sustrato auténtico de una ciencia de las "almas"" (Husserl 1976a, 216).

Antes de abordar plenamente la tarea de establecer este subsuelo primitivo, importa precisar que el camino que propone seguir la filosofía transcendental de Husserl para investigar el ser psíquico, no es el que se ha perfilado en las páginas anteriores, sino otro radicalmente diferente. El planteamiento de Fodor acerca de las relaciones entre psicología popular y psicología cognitiva coincide plenamente con el planteamiento que Husserl caracteriza genéricamente como propio de la filosofía objetivista: partir del suelo del mundo (psíquico) previamente dado en la psicología popular, y remontarse paulatinamente hacia la investigación científico-natural de las propiedades "reales" de ese dominio de ser peculiar. Cabe, con todo, y siempre según Husserl, realizar una investigación del ser psíquico que no comparta esta estrategia general de carácter científico-natural. Esta segunda vía es la que quiere seguir propiamente la filosofía transcendental que, en este sentido, como en tantos otros, se contrapone radicalmente a la filosofía objetivista. El suelo absoluto del que parte la filosofía transcendental es el de la "subjetividad absoluta, transcendental". Según la reconstrucción que hace Husserl de la historia de la filosofía a partir de la modernidad, el desarrollo de las ideas y de los distintos sistemas filosóficos occidentales se puede considerar como un debate prolongado y apasionado entre estas dos

³ Esta relativa equiparación provisional entre la psicología popular y el mundo de la vida no debe dar lugar a pensar que identifico plenamente los contenidos de ambas nociones. La equiparación sólo se hace desde el punto de vista de cómo caracterizamos al ser psíquico antes de la construcción de cualquier ciencia que lo tome como su objeto propio.

ideas de filosofía. En el centro de este debate y como "campo de las decisiones" ontológicas acerca del ser psíquico sitúa también Husserl a la psicología.

Ahora bien, ¿en qué sentido y con qué finalidad cabe otorgar a la psicología esta función privilegiada? Creo que para contestar a esta pregunta se pueden adoptar dos estrategias diferentes. La primera la señala de pasada el propio Husserl. Consistiría en forzar todo lo posible los supuestos implícitos de una psicología que se presentara como psicología científico-objetiva y que tomara como punto de partida las propiedades del ser psíquico que la psicología de sentido común acepta como evidencias originarias acerca del alma. Esta estrategia abunda en la idea de que la psicología científica debe ser el árbitro de la ontología de lo mental.

<Un> camino hacia una filosofía transcendental tiene que poder trazarse sobre una psicología elaborada concretamente (Husserl 1976a, 210).

La segunda estrategia, la seguida propiamente por Husserl, consistiría en elaborar una psicología "con otra actitud" que la que podría desarrollarse desde la actitud científico-natural. Sin embargo, desde la actitud transcendental--así llama Husserl a esta otra actitud--la psicología dejaría de tener los mismos planteamientos de tareas y de métodos que la psicología científico natural. La gran tarea de la psicología científico-natural es el alma, la mente. La tarea primordial de la psicología transcendental es el Yo, la subjetividad transcendental.

Si se dejan al margen los detalles sobre las difíciles relaciones entre la actitud transcendental y la actitud natural--Husserl comprende en cierta ocasión "la actitud natural unilateralmente cerrada como una actitud transcendental peculiar" (1976a, 209)--la tarea general de este libro consistirá en seguir la primera estrategia para recorrerla hasta sus últimos problemas

de fundamento. Si hay que recorrer el camino de la fenomenología trascendental que Husserl ha abierto e indicado, es bueno convencerse antes de que el más obvio, el de la filosofía objetivista, conduce a problemas y paradojas irresolubles. Esta tarea marca las directrices de mi trabajo que, en este sentido, está inspirado en las que guiaron a Merleau-Ponty cuando este insigne fenomenólogo escribió *La estructura del comportamiento*: "nuestro objetivo es comprender las relaciones entre la conciencia y la naturaleza" (1953, 19).

Sea una u otra estrategia la que se precise seguir en el inicio mismo o a la larga, ambas necesitan comenzar, en cualquier caso, con la caracterización del "último sustrato auténtico de la ciencia de las almas".

De acuerdo con Husserl, mi "imperar" yoico sobre mi cuerpo vivo (*Leib*) y la intencionalidad, son los dos datos esenciales que se pueden extraer de la descripción ametafísica de la experiencia mundano-vital del ser psíquico. En la experiencia fenomenológica, mi cuerpo no lo vivo originariamente como un simple cuerpo físico (*Körper*) entre los otros muchos que componen el sistema natural cerrado de cuerpos que pueblan el universo natural, sino como cuerpo vivo (*Leib*) sobre el que yo "impero" o "mando" (*ich walte*); un cuerpo al que "animo", al que "muevo". Sin embargo, pronto descubro también que no solamente impero sobre mi cuerpo, sino que él es—dentro de ciertas limitaciones conocidas intuitivamente por todos—un instrumento que colabora conmigo y con mis intenciones de una manera más o menos armónica y eficaz cuando intento "imperar" también sobre el mundo físico circundante: yo cojo, atraigo, subo o bajo este o aquel objeto (Husserl 1976a, 216). En general un "yo puedo hacer". Es así como experimento de forma originaria en mi cuerpo vivo la "corporeización de las almas" (Husserl 1976a, 220). Sólo cada uno en sí mismo puede encontrar propiamente esta clase de experiencia única.

Esta primera descripción sumamente genérica da lugar inmediatamente a una serie de preguntas que constituyen motivaciones esenciales para emprender la investigación psicológica científica. A este respecto, leemos en Husserl lo siguiente:

El preguntar reflexivo originario se orienta ahora a lo siguiente: qué y cómo son las almas--en primer lugar las almas humanas--en el mundo, en el mundo de la vida; así, pues, cómo "animan" cuerpos vivos, cómo están localizadas en la espacio-temporalidad, cómo "vive" anímicamente cada una de ellas en tanto que tiene "conciencia" del mundo en el que vive y es consciente de vivir; cómo experimenta cada alma su cuerpo.... (Husserl 1976a, 215-216).

Con esta primera descripción a la vista y con las preguntas a que dan lugar en primera instancia, se puede formular el segundo dato esencial de la experiencia mundano-vital de lo psíquico que destaca Husserl. Este no es otro que el de la intencionalidad, que se manifiesta de múltiples formas en el mundo vital: el lenguaje, la percepción, la persecución de fines y medios, etc. (Husserl 1976a, 236).

Podría parecer que estos dos datos esenciales coinciden en sus líneas generales con las propiedades del ser psíquico que se pueden desprender de la queja que Sócrates expresaba a propósito de las explicaciones mecanicistas de su conducta. Lo que estaría reclamando Sócrates sería, en el lenguaje de Husserl, su imperar yoico sobre su cuerpo vivo; imperar en el que se manifestarían sus cogitaciones y sus disposiciones intencionales en general, en virtud de la docilidad y transparencia--una vez más dentro de ciertos límites--con las que su cuerpo vivo "traduce" públicamente estas intenciones subyacentes. Sin embargo, si se expresa de esta manera la coincidencia general de las propiedades que Sócrates y Husserl atribuyen al ser psíquico de la experiencia común, se deja escapar una diferencia fundamental. Donde Sócrates hablaba explícitamente de que sus cogitaciones son las "verdaderas

causas" de por qué está en la cárcel, Husserl, quizá más cauto y, en cualquier caso, mucho más receloso, habla de un vago "imperar".

Las razones de esta retracción se darán más adelante. Pero, sin duda alguna, ya se puede decir con antelación que una de las motivaciones fundamentales de esta timidez filosófica es que hablar aquí de "causalidad intencional" complicaría excesivamente la investigación psicológica, al tiempo que pondría en entredicho el método exclusivamente descriptivo con el que la fenomenología trascendental--la verdadera psicología, según Husserl--, encara la investigación del ser psíquico⁴. El problema que se presenta como más espinoso cuando se desea, por considerarlo inevitable, seguir el planteamiento de Sócrates y de Fodor, es el de dar con una buena idea que nos indique con precisión cómo algo puede entrar en relaciones causales preservando, sin embargo, relaciones semánticas básicas.

Sin embargo, antes de ahondar aún más en estas motivaciones, es importante subrayar el peso que otorga Husserl a los dos datos originarios ya mencionados para denunciar el desenfoque radical de los planteamientos, fines y métodos de la psicología racional transmitida a partir de los supuestos dualistas de Descartes. La denuncia de Husserl se concreta en esta tesis:

En lugar de ello <en lugar de empezar con la caracterización del subsuelo de experiencias originarias del que ha de partir la investigación> la psicología no comenzó en modo alguno con un

⁴ Por otra parte, si la fenomenología trascendental es la verdadera teoría del conocimiento, la psicología y la teoría del conocimiento son dos nombres para el mismo género de investigaciones. Por supuesto, la psicología de la que se habla aquí es la psicología fenomenológico-descriptiva. Por otra parte, este modo indiferente de hablar de la psicología y de la teoría del conocimiento se puede encontrar también en las lecciones de Ortega y Gasset que se publicaron póstumamente con el título de *Investigaciones psicológicas*. Hoy día, sin embargo, existe la convicción no tanto de que la psicología sea la verdadera teoría del conocimiento, sino de que "la filosofía de la mente es la nueva forma de la tradicional teoría del conocimiento." (Debo esta observación a una comunicación personal del profesor Pascual Martínez Freire. (Véase también (Martínez-Freire 1993, 38).)

concepto de alma creado originariamente, sino con un concepto que proviene del dualismo cartesiano, un concepto que venía dado por una idea constructiva ya precedente de una naturaleza corpórea y de una ciencia natural matemática. (Husserl 1976a, 216).

Una vez conocidas las raíces de la distorsión que introdujo el cartesianismo en el concepto originario de alma, concepto que debe difundir sentido a la investigación psicológica especializada, es hora de adentrarse en las motivaciones profundas que impulsaron a Husserl a hablar de un mero "imperar", pero no de un "causar". Los tres frentes que guían ahora la discusión son la espacio-temporalidad, la causalidad y la individuación de lo propiamente psíquico. Estos tres conceptos están estrechamente relacionados. Con todo, convendría tratar en primer lugar la espacio-temporalidad en unión con la causalidad. Según Husserl, la espacio-temporalidad es "la forma universal del mundo real" y, en primera instancia, de los cuerpos (Körper) (Husserl 1976a: 220). La doctrina clásica afirma que lo que es material es, *eo ipso*, espacio-temporal y tiene capacidad causal (eficiente). Por consiguiente, si se admitiese que los estados intencionales que Sócrates se adscribe y presenta como las verdaderas causas de su comportamiento fueran realmente causas, entonces, muy probablemente, habría que caracterizarlos como materiales y, se sigue de suyo, como espacio-temporales. Cualquier cosa que tiene capacidad causal es material y, *eo ipso*, espacio-temporal. Husserl se pregunta ahora si las almas poseen, al igual que los cuerpos y en el mismo sentido, espacio-temporalidad. Acogiéndose a la doctrina de su maestro Brentano, Husserl enuncia simplemente que "el ser anímico en y para sí no tiene ninguna extensión ni ubicación espacial" (1976a, 220). La conclusión cae por su propio peso: lo que no es espacio-temporal, no es material y no tiene, por tanto, capacidad causal. Un yo, propiamente, no causa "por medio" de sus actitudes proposicionales. Antes bien, la argumentación precedente obliga

incondicionalmente a diluir poco a poco la forma de expresión que utiliza Sócrates, que es ontológicamente muy arriesgada y, quizá, apresurada, hasta llegar a un vago "imperar yoico". El problema que surge de inmediato es entender cómo puede realizarse este "imperar" yoico, cómo puede explicarse más allá de estas caracterizaciones intuitivas que se establecen sobre el sustrato descriptivo básico que ofrece el análisis "desprejuiciado" del mundo de la vida. Que alguna clase de dificultad esencial debe estar contenida ya en estas descripciones, puede colegirse del hecho de que Husserl escriba siempre la palabra "imperar" entre comillas, como si su sentido remitiese a dificultades encerradas más allá de las connotaciones inmediatas que aquella palabra tiene. ¿Hacia dónde remitiría? Creo que al problema de la inteligencia encarnada planteado en toda su crudeza y radicalidad. Pero no sólo a él. También forman parte de estas dificultades los problemas relativos a los criterios que se pueden arbitrar para establecer cuándo un sistema material, un cuerpo (en el sentido de Körper), se debe determinar como un cuerpo animado yoicamente (Leib). De acuerdo con las descripciones preliminares, el único cuerpo del que puedo tener evidencia originaria de estar "animado" por un yo ejecutivo, es mi propio cuerpo vivo. La certeza con la que predique un yo imperante como "parte animadora", por así decir, de otro cuerpo que no sea el mío, ha de ser necesariamente de un grado inferior, probabilístico, que se alcanzaría por medio de inferencias a partir de su conducta pública observable.

La conducta de los demás es contingente con respecto a determinados estados, procesos y eventos mentales: habiendo advertido que nuestra propia conducta tiene tales y tales antecedentes, y que, por otro lado, es muy similar a la que producen otras personas, llegamos a la hipótesis de que en los demás deben darse operaciones mentales semejantes a las que se dan en nosotros. (Fodor 1980b, 129).

Sócrates tenía la certeza de que él "imperaba" sobre su cuerpo vivo y, por esta razón, renegaba de que nadie explicara las causas antecedentes de su conducta de una forma mecanicista. Esta certeza que él reclamaba para sí estamos dispuestos a concedérsela sobre la base de nuestra propia experiencia. Si yo tengo experiencia de que impero sobre mi cuerpo vivo, de que él es el "órgano" que coopera conmigo--más o menos dócilmente--en mi actuar sobre los objetos del mundo y también en mi padecerlos; si mis intenciones y, en general, los fenómenos intencionales que me adscribo, tienen una "traducción" armónica en mis movimientos corporales como sus instrumentos fiables de ejecución, entonces estoy dispuesto a conceder que los movimientos corporales de Sócrates también son "traducciones" de todo género de fenómenos intencionales de los que no tengo, ni puedo tener, una experiencia sensible directa. Como se suele decir normalmente, no puedo observarlos. Pero admito que él se adscriba estos fenómenos para dar razón de su conducta porque yo me siento fuertemente inclinado a adscribírselos también sobre la base indirecta de cómo puedo dar sentido a lo que él está haciendo. Pronto me convenzo, sin embargo, de que Sócrates, a quien considero ya en este momento como otro sujeto distinto que yo, estaría dispuesto a aplicar la misma clase de razonamiento inductivo conmigo que yo le he aplicado a él. Es así como las explicaciones de la psicología de sentido común se van difundiendo entre los sujetos y los cuerpos vivos y como se van atrincherando cada vez más honda e inexpugnablemente en nuestra manera de comprender nuestro actuar en el mundo y el de nuestros semejantes. El siguiente paso es que yo, como sustrato último de la atribución de actitudes proposicionales por parte de los demás, estoy tan alejado de ser y aparecer "transparente" a los demás en mi obrar intencional como ellos lo están para mí. En este sentido, puedo estar máximamente interesado en comprometerme a buscar con los otros criterios fiables para "reconocer yoes" actuantes. La tarea teórica que

aquí se anuncia es, como puede comprenderse, de una importancia suprema. Esta búsqueda está llena, sin embargo, de problemas y de grandes peligros. Uno de los peligros fundamentales, como se irá haciendo cada vez más claro a lo largo de la investigación, es que en el camino de esta búsqueda tiene que darse una mezcla inevitable entre planteamientos en primera y en tercera persona.

El punto de vista de la primera persona y los análisis descriptivos a él pertenecientes, son exclusivamente fenomenológico-intencionales. El de la tercera persona es el punto de vista heterofenomenológico-causal⁵ que sólo puede expresar "el que ya sabe". Sin duda, este último enfoque ha sido el más explotado por la literatura anglosajona dedicada a estos problemas filosóficos. Sin embargo, el resultado de esta segunda vía es que, a la larga, se iría perdiendo, en un sentido muy profundo, lo que de incuestionablemente yoico aparecía en el comienzo de las investigaciones. Este es un aspecto esencial que debería abordar toda filosofía de la psicología pero que, sin embargo, ha estado usualmente alejado de las discusiones críticas sobre los fundamentos y consecuencias de la psicología cognitiva. El problema del yo y de la conciencia no ha asomado realmente--más allá de tomar su nombre en vano--cuando se ha tratado de profundizar en el problema de la intencionalidad desde este punto de vista. Recientemente, sólo Searle, entre los filósofos de la "tradición analítica", ha tratado de irritar tímidamente esta herida abierta y nunca atendida hasta sus raíces últimas por los filósofos de la mente contemporáneos, quienes, en general, no la han considerado especialmente molesta por razones esenciales. Pero lo es y mucho.

En resumen, los dos datos esenciales que, según Husserl, delimitan y caracterizan el sustrato último de la investigación psicológica-imperar

⁵ La voz "heterofenomenología" está tomada de Daniel Dennett (1982a, 153). Dennett contrapone la autofenomenología a la heterofenomenología.

yoico e intencionalidad--llevan a plantearse qué quiere decir propiamente "imperar yoico" y a qué esfera de problemas remite genéricamente esta última noción. Estos problemas son: 1) cómo es posible la realización efectiva de este imperar yoico y 2) cuáles son los criterios para reconocer yoes actuantes semejantes a mí. Pero esta investigación, debido simultáneamente a la conciencia que tengo de que los otros yoes no son "transparentes" a mi interés por conocerlos en tanto que sistemas intencionales, y de que yo, correspondientemente, no lo soy tampoco para ellos, se complica hasta extremos insospechados por la razón de que la base con la que cuento y cuentan los demás para su prosecución, es la conducta observable públicamente y las inducciones que se generan paso a paso a partir de ella. Así, pues, cuando comienzo esta tarea teórica, tengo que estar dispuesto a irme dejando perder poco a poco como yo para convertirme intersubjetivamente en tema de una investigación que tiene cada vez más el carácter de la investigación objetivo-natural. Los problemas filosóficos de la fundamentación de las ciencias cognitivas tienen el destino peculiar--de ahí su enorme dificultad--de tener siempre un ojo puesto en el punto de partida esencialmente yoico del mundo de la vida y que difunde sentido una y otra vez a la investigación psicológica científico-natural, y otro ojo puesto en las exigencias y compromisos ontológicos que la realización concreta de esta investigación obliga a contraer. Este punto de vista bifocal es una fuente constante de tensiones y de problemas para todo psicólogo que tenga preocupaciones filosóficas y que no se conforme con la "ingenuidad" de una práctica científica que funciona automáticamente. Pero también lo es para todos los filósofos que estén empeñados en no querer descender hasta los problemas "menores" de la investigación psicológica natural y de sus supuestos por considerar que el área en la que hay que debatir los problemas que se encierran bajo el título "ego" no

tienen nada, o muy poco, que ver con la forma científico-objetiva de hacer psicología.

Después de estas aclaraciones generales, pero necesarias, acerca de la multitud de tareas que se originan y entrecruzan cuando se quiere investigar el imperar yoico, importa retomar el problema de la "causalidad" de lo psíquico, pero esta vez en unión con el problema de su individuación y no con el de su espacio-temporalidad. Su comprensión es únicamente posible cuando se contraponen la causalidad e individuación de los cuerpos y la causalidad e individuación de las "psiques". Lo primero que apunta Husserl es que *el sentido* en que se habla de la "causalidad" de lo psíquico es completamente diferente del sentido que se aplica a lo corporal espacio-temporal. Entender propiamente cuál es el matiz que distingue entre estos dos sentidos supone una tarea sumamente ardua.

Un cuerpo es lo que es en tanto que este cuerpo determinado: un sustrato de cualidades "causales" localizado espacio-temporalmente en su propia esencia. Así, pues, si se hace desaparecer la causalidad, el cuerpo pierde su sentido de ser como cuerpo, su identificabilidad y diferenciabilidad como individualidad física. Pero el Yo es "éste", y posee la individualidad en sí y a partir de sí mismo, no a partir de la causalidad. (Husserl 1976a, 221-222).

La dificultad inherente a esta cuestión de las relaciones entre la causalidad y la individuación es semejante a la que aparecía anteriormente cuando se discutió el problema de la espacio-temporalidad del ser psíquico. Desde el punto de vista originario en el que yo me percibo a mí mismo, no me sentiría cómodo si alguien estuviera dispuesto a defender respecto de mí que mi *haecceitas* depende esencialmente de las interacciones causales entre "el mundo" y mis estados anímicos, o entre mis estados anímicos entre sí, o entre éstos y mis movimientos corporales. Parece que yo soy una entidad individual a la que no hace la más mínima contribución el

hecho de estar impropia e indirectamente localizada en un cuerpo, y como "participando" de su espacio-temporalidad (Husserl 1976a, 220).

En tanto que tal <el Yo> tiene de antemano en sí su unicidad. Para el Yo, tiempo y espacio no son principios de individuación; el Yo no conoce ninguna causalidad natural, la cual, según su sentido, es inseparable de la espacio-temporalidad; su actuar es un imperar yoico (*ichliches Walten*) y éste acontece inmediatamente por medio de sus cinestesias en tanto que imperar en su cuerpo vivo, y sólo mediatamente (puesto que éste también es cuerpo) sobre otros cuerpos. (Husserl 1976a, 222).

(b) Las cinestesias: ¿La glándula pineal de Husserl?

La importancia que tiene el texto que se acaba de citar es que delega momentáneamente la discusión de la individuación del yo--nunca sujeta a la causalidad natural--al problema de su actuación. Su actuar no es un causar sino un "imperar". Sin embargo, esta comprensión pasaría necesariamente por la comprensión de las cinestesias como el lugar privilegiado en el que se realiza y manifiesta esta capacidad ejecutiva del yo⁶.

Mi tesis es que es posible que haya una precomprensión oscura de lo que sea este actuar inmediato del "imperar" yoico sobre su cuerpo vivo. Pero no creo que haya una comprensión propia de cuáles sean sus condiciones de posibilidad. Varios son los problemas que se dan aquí la mano. El primero es la sospecha de que yo no "muevo" mi mano de la

⁶ Sin embargo, la discusión sobre las cinestesias es relativamente independiente de problemas tan interesantes como el de la individuación del yo. Cfr. Husserl, Edmund. "Seefelder Manuskripte über Individuation." En *Zur Phänomenologie des inneren Zeitbewusstseins (1893-1917)*. (Ed. Rudolf Boehm. Husserliana X. The Hague: Martinus Nijhoff), 1966: 237-268. Una idea de las dimensiones metafísicas que llegó a adquirir el problema de la individuación en la filosofía de Husserl, puede encontrarse en: Husserl, Edmund. *Zur Phänomenologie der Intersubjektivität. Texte aus dem Nachlass. Dritter Teil: 1929-1935*. (Ed. Iso Kern. Husserliana XV. The Hague: Martinus Nijhoff), 1973: 666-670.

misma manera que una bola de billar mueve a otra cuando choca con ella. Sin embargo, mi mano--aquello sin lo cual mi "imperar" se quedaría en mera posibilidad impotente de actuar mediatamente sobre otros cuerpos--sí que puede mover este o aquel objeto según el modelo de las bolas de billar. Si el modelo intuitivo de las bolas de billar es el de la causalidad natural, entonces no quiero decir que mi "imperar" yoico sea un causar natural. La contrapartida de este recelo es relegar al irracionalismo cómo es que el Yo dispone libremente de su súbdito corporal.

Recurrir a las cinestesias, como hace Husserl, no parece que sea de gran ayuda. Las cinestesias son el punto misterioso donde lo yoico y lo físico se dan la mano. Argumentaré que si no se puede solucionar el problema de las cinestesias sin hablar de cómo los estados intencionales del sujeto son causa (sea cual sea el sentido en que haya que hablar de causa aquí), entonces será difícil no reconocer que los dos datos inmediatos que la psicología popular provee como el punto de partida inteligible de la investigación psicológica son la causalidad y la intencionalidad de los estados representacionales del sujeto.

Antes, como es obvio, conviene recorrer algunos pasos fundamentales para justificar esta aserción. En *Experiencia y Juicio* las cinestesias son conceptualizadas como movimientos corporales que están, ante todo, al servicio de los procesos perceptivos del sujeto y, en concreto, al servicio de la percepción externa (Husserl 1985, 89). Gracias a estos movimientos corporales, el objeto de percepción externa va mostrando nuevos lados hasta ahora ocultos para el sujeto. Su rendimiento también puede ser el de hacer que un objeto que no se nos está dando en el modo de la percepción, pueda dárseos él mismo en persona. Cuando doblo una esquina percibo objetos antes ocultos para mí.

Los movimientos cinestésicos son, por otra parte, efectos o resultados (*Auswirkungen*) "de las tendencias de la percepción", tendencias que se

manifiestan primariamente en la orientación no simplemente puntual del yo hacia el objeto. Estos movimientos corporales tienen la función esencial de desarrollar estas tendencias perceptivas. Ellos proveen al sujeto, por así decir, con nuevas imágenes (Bilder) del objeto percibido. El objeto de percepción externa se da primariamente en una imagen tras la primera orientación del yo, y llega a constituirse paulatina y reglamentadamente para el sujeto mediante la síntesis de nuevas imágenes que son vividas por aquél como imágenes del mismo objeto. Entre otras condiciones, este constituirse gradual y reglamentadamente el objeto para el sujeto se realiza precisamente gracias a las cinestesias. Los movimientos del cuerpo vivo (Leibesbewegungen) y los modos de aparición del objeto correrían paralelos (Rang 1973, 161).

En estos movimientos corporales, yo, como "sujeto cinestésico" (Rang 1973, 162), me manifiesto a la vez como productivo y receptivo respecto del objeto. En ciertas condiciones, está en mi mano interrumpir la secuencia de imágenes que se produce constantemente de una forma paulatina cuando dejo que las cinestesias entren en juego libremente. Si cierro mis ojos o me niego a girar la cabeza para ver el objeto que, en su movimiento, se ha situado detrás de mí, me mantengo en una relación activa respecto de las pretensiones del objeto, de las tendencias perceptivas que ha originado en mí. Yo interrumpo aquellas cinestesias que caen dentro de mi radio de acción; impero" sobre ellas; soy activo. Por el contrario, no está en mi poder que no se originen nuevas y nuevas imágenes del objeto si "dejo que se produzcan las cinestesias....<Si> pongo en juego estas y aquellas cinestesias frente al objeto, se presentarán estas y aquellas imágenes" de grado o por fuerza (Husserl 1985, 89).

La argumentación comienza ahora. Si una de las funciones esenciales de los movimientos corporales cinestésicos, si no la principal, es estar al servicio de la percepción externa de objetos, entonces esta función sería un

buen criterio para clasificar un movimiento corporal específico como una cinestesia. Si esto es así, entonces el ámbito de manifestación de las cinestesias debería comprender tanto los movimientos imperceptibles de convergencia y acomodación oculares—los movimientos sacádicos—, como movimientos de ojos más amplios o movimientos de cabeza. Sin embargo, hay que advertir que Husserl sólo incluye en sus ejemplos de cinestesias aquellos movimientos del cuerpo que, en caso de necesidad, el sujeto podría controlar voluntariamente. Así, pues, cinestesias serían en primer lugar movimientos de cabeza o de tronco, pero no movimientos de acomodación y enfoque oculares. De todas las maneras es preciso insistir en que debería incluir ambas clases de movimientos si se atuviese a la función principal que les asigna: proveer nuevas imágenes del objeto percibido. Que los movimientos sacádicos sean movimientos "objetivados", movimientos de los que el sujeto no tiene un conocimiento y un control consciente directo no debería ser ningún obstáculo especial a la hora de hacer un inventario de todos los movimientos que contribuyen esencialmente a la percepción de los objetos, sea como sea que tenemos conocimiento de estos movimientos.

De todas formas, Husserl piensa que las cinestesias, aun poseyendo el carácter de procesos subjetivos activos, no son, o al menos no necesariamente, acciones voluntarias (Husserl 1985, 89-91). Involuntariamente muevo mis ojos o mi cabeza sin atender primariamente a ellos cuando me dejo arrastar por las tendencias perceptivas que el objeto ha encendido en mí. Es más, algunos movimientos cinestésicos—si es que de verdad hay que incluir todos los que se han mencionado anteriormente—pueden pasar desapercibidos para el sujeto y, en este sentido, fuera de su control voluntario. Así, los movimientos de enfoque ocular. Sin embargo, de acuerdo con Husserl, ¿habría que suponer que estos movimientos seguirían siendo procesos subjetivos activos?

El obstáculo fundamental, sin embargo, no radica primariamente en el carácter voluntario o involuntario de los movimientos cinestésicos—aunque también en él—, sino en cómo los clasificamos. Para clasificar toda esta variedad de movimientos corporales como cinestesias lo esencial es que ellos hayan servido funcionalmente para la obtención de nuevas "imágenes" del objeto percibido. No muevo mi tronco y mi cabeza para hacer cierto ejercicio de gimnasia o para desentumecer mis músculos, sino para ver mejor en todas sus fases este objeto que se mueve a mi alrededor. Los movimientos corporales se clasifican transversalmente por las intenciones del sujeto y no por su descripción física. Este movimiento corporal concreto es una cinestesia porque sirve a las intenciones implícitas que consisten en ir desarrollando las tendencias perceptivas que el objeto ha encendido en mí y a las que yo respondo con una orientación que no es meramente instantánea. El mismo movimiento corporal es un movimiento gimnástico porque sirve a las intenciones de seguir las reglas de esta tabla gimnástica. Una teoría de las cinestesias como conducta perceptiva de un sujeto cinestésico no puede ser sustituida por una teoría de los movimientos corporales. El comportamiento y, por consiguiente, también el "comportamiento" perceptual, consiste en acciones, y las acciones clasifican transversalmente los movimientos. El mismo movimiento corporal puede ser una cinestesia o un movimiento gimnástico.

A la luz de estas descripciones de las cinestesias, ¿dónde queda el imperar inmediato del yo sobre su cuerpo vivo? ¿No está detrás del proyecto de Husserl hablar de los movimientos cinestésicos como si éstos pudieran producirse fuera de todo control voluntario y, por tanto, de toda actividad auténtica del yo? Si esto es así, ¿en qué sentido se puede decir que en los movimientos de acoplamiento y enfoque ocular, por ejemplo, hay una actividad inmediata del yo sobre su cuerpo vivo? Y respecto de los movimientos cinestésicos que sí pueden caer dentro del control

voluntario del sujeto, ¿no se plantean para ellos las mismas preguntas que se querían resolver con su introducción? ¿No es legítimo preguntar cómo es posible que mi cuerpo ande alrededor de la pieza de museo cuando la observo por nuevos lados? En este caso, los movimientos cinestésicos están guiados por un interés, por una intención, que, como ya se anunció, es necesario tener en cuenta cuando se quieren clasificar determinados movimientos corporales como cinestesias y no como movimientos gimnásticos. Pero el problema de Sócrates es precisamente saber cómo las intenciones del sujeto pueden guiar los movimientos corporales; cómo las verdaderas causas están en armonía, por así decir, con los estados corporales. Por consiguiente, nada o muy poco se ha avanzado por este camino. El imperar del yo sigue siendo un misterio, y el problema de la "inteligencia encarnada" sigue sin resolver. Decir que el imperar yoico acontece inmediatamente por medio de sus cinestesias no soluciona el problema de Sócrates porque no sabemos cómo tiene lugar este "acontecer", ni cuáles son sus condiciones de posibilidad.

Husserl acusaba, con razón, al dualismo cartesiano de querer imponer una concepción de lo anímico y de su método de investigación que no partía de bases descriptivas adecuadas. El cartesianismo imponía una "idea constructiva ya precedente" acerca de lo que es el ser psíquico (1976a, 216). Sin embargo, Husserl corre el mismo peligro. El imperar yoico no es un causar basado en el modelo de las bolas de billar, pero no se entiende qué es y cuáles son sus condiciones de posibilidad.

¿Sócrates o Husserl? ¿Intencionalidad y causalidad como propiedades del ente psíquico tal y como se nos da éste en el mundo de la vida, o intencionalidad e imperar yoico? Creo que Sócrates tenía razón.

& 4. Datos originarios y construcción teórica.

La precomprensión del ser psíquico realizada sobre el sustrato originario del mundo de la vida (de la psicología popular), es ella misma una fuente inagotable de interrogantes profundos y apasionantes. Estos interrogantes constituyen la motivación primaria inteligible que ha de conducir a investigaciones psicológicas de cuño científico-positivo. Sin embargo, esta precomprensión no se realiza con criterios y descripciones unívocos. Causar o "imperar" se llama el primer dilema. He tratado de argumentar, sin embargo, que estamos abocados a hablar de causación y que, por lo tanto, el punto de partida inteligible de la psicología científica habrá de ser la causalidad y la intencionalidad de las actitudes proposicionales. Sobre estos dos datos descriptivos primarios del ser psíquico se alza el abrumador problema del yo y de la conciencia. El imperar yoico es un enigma. Sin embargo, la estrategia para solucionarlo--si es que se puede solucionar--no consiste en eliminar la parte enigmática que corresponde al yo, sino en explicar la parte enigmática que corresponde a su imperar. Por desgracia, la tarea reflexiva que bajo el título de "filosofía de la mente" investiga sistemáticamente los fundamentos de la psicología cognitiva ha considerado, en general, los problemas que giran en torno al yo y a la conciencia como derivados, cuando no los ha ignorado resueltamente. Por el contrario, se ha centrado primordialmente en el problema del imperar del yo bajo la forma del problema de la inteligencia encarnada.

Este peculiar viraje que ha tomado la investigación en filosofía de la mente, ha tenido como consecuencia que el problema de la intencionalidad se plantee independientemente del problema del yo y de la conciencia. Las actitudes proposicionales tienen propiedades intencionales y causales desde el punto de vista de la psicología popular. Centrarse en el

problema de la inteligencia encarnada ha dado lugar a que los investigadores intentaran buscar una idea con la que se pudiera "coser" la intencionalidad con la causalidad de las actitudes proposicionales. Pero esta idea ha vuelto la espalda al yo y a la conciencia. El resultado final ha sido una teoría de la intencionalidad de las actitudes proposicionales al margen de la conciencia. Sin embargo, quiero subrayar que este viraje no es gratuito. Antes bien, obedece al punto de partida de la investigación psicológica tal y como se ha establecido en los párrafos precedentes. Explicar el isomorfismo entre las relaciones causales y semánticas de las actitudes proposicionales se ofrece como un reto ante nuestros ojos.

Según estas consideraciones, no me atrevo a decir directamente que en todo este enjambre de problemas se oculte una paradoja de difícil solución. Sólo diré que el problema de la inteligencia encarnada empuja en una dirección que no se aviene muy bien con la dirección hacia la que quiere tirar el dato, también primario, de la conciencia que tenemos de nuestras actitudes proposicionales y de sus objetos.

A este primer foco de perplejidades se une el carácter intersubjetivo que alcanza finalmente la prosecución de mi comprensión compartida de lo que supone ser un yo y de los criterios para reconocerlo. Este estudio en tercera persona se une al estudio, también en tercera persona, de las condiciones de posibilidad del imperar yoico. Aquí se abren dos frentes. Uno en primera persona: yo anulo cualquier caracterización de mi imperar sobre los órganos de mi cuerpo que quiera seguir el modelo de las bolas de billar. El otro en tercera persona: investigación de las condiciones de posibilidad de mi imperar yoico y de los criterios intersubjetivos para reconocer cualquier yo ejecutivo que no sea el mío. Si digo en mi interior "yo nuevo", entonces no creo entender nada semejante a una bola de billar que empuja mi mano o mis piernas. Desde el punto de vista de la primera persona me prohíbo hablar de mí como causa natural. En este

sentido soy una "unidad indescomponible". El análisis de los movimientos de mis órganos corporales y su posterior ensamblaje armónico desde "el exterior", no pueden reconstruir mi yo. Pero mi acceso a cualquier otro yo ajeno al mío tiene que ser una vía que se vaya construyendo paso a paso con la ayuda de inducciones y analogías que son de gran ayuda y sumamente predictivas⁷. Si yo me adscribo actitudes proposicionales y creo firmemente ver en mis conductas intencionales de todo tipo el "reflejo" en el que mis actitudes proposicionales se manifiestan intersubjetivamente, esto es, públicamente, entonces, para poder predicar "es un yo" de cualquier otro, me sirvo de una analogía muy primitiva e íntimamente arraigada en mi forma de ver el mundo--si se trata realmente de una analogía--gracias a la cual conductas análogas a las mías deben de ser también reflejos de otra actividad yoica distinta de la mía pero análoga a ella. Una vez más, desde el momento mismo en el que "constituyo" otro yo, me veo envuelto irremediablemente en una visión del mundo en la que yo soy un nodo más en una red de relaciones yoicas. Cuando constituyo un yo distinto del mío, simplemente soy, de pronto, otro yo más que nadie puede "ver" a no ser que aplique el mismo razonamiento que acabo de hacer conmigo. Aquí se produce un viraje radical en la forma de verme a mí mismo: no sólo soy un yo, sino un nodo que, seguramente, es tan opaco en lo que tiene de yo como lo son los otros yoes para mí. En tanto que nodo en un red de relaciones y de deducciones, puedo ser tema de una investigación intersubjetiva, una investigación en tercera persona a la que, por mucho que insista en decir "yo", junto con sus modalidades "yo nuevo", "yo levanto", "yo percibo" etc., no puedo sustraerme sin caer en el solipsismo. Me comprometo a buscar intersubjetivamente formas de reconocimiento de

⁷ Sin embargo, no es obvio que el paso a la intersubjetividad haya de ser secundario y se dé en estas etapas que son, por otra parte, las etapas en las que más ha insistido gran parte de la filosofía de la mente contemporánea.

otros yoes más allá de cualquier precomprensión intuitiva, pero borrosa, de lo que esto supone. Solamente así podré estar seguro de que nadie puede confundirme con otra cosa que no sea un yo. Pero esta tarea es una tarea en tercera persona. Irremediablemente lo es. Es una tarea en la que me voy perdiendo como yo. Eso significa radicalmente mundanizarme.

De esta investigación en tercera persona también forma parte el intento de arrojar alguna luz sobre la legitimidad con la que la psicología popular persiste en caracterizar a las actitudes proposicionales tanto semántica, como causalmente.

Antes de terminar esta sección importa hacer un pequeño apunte. Aunque en la literatura filosófica contemporánea todos estos interrogantes han quedado circunscritos, casi exclusivamente, dentro del límite demarcado por los debates de la filosofía de la mente, o filosofía de la psicología, su trascendencia e implicaciones para otras áreas del saber filosófico, como la antropología filosófica, por ejemplo, me parecen ser de todo punto irrevocables. El modelo que domina desde hace tres decenios la explicación psicológica científica de los procesos cognitivos superiores--el modelo de la mente como un sistema de procesamiento de la información--, concibe que las actividades intencionales inteligentes de todo tipo que realiza el hombre, se pueden simular por medio de modelos computacionales efectivamente realizables en sustratos materiales ilimitadamente variables. Como dice el *motto* tan querido a los investigadores cognitivos de orientación funcionalista, se puede hacer inteligencia casi a partir de cualquier cosa. Todo es cuestión de tener la organización funcional adecuada. Es una cuestión de diseño arquitectónico y de programación, y no de los materiales de construcción que se empleen. Si este proyecto no se puede realizar ahora mismo--así se suele argumentar--, es cuestión de tener un poco de paciencia y esperar el tiempo necesario hasta conseguirlo. Excepto la esfera de las emociones y de los sentimientos, que se sustrae a sí

misma, por razones esenciales, a la investigación psicológica científico-cognitiva--véase cualquier libro de texto de psicología cognitiva para constatar este hecho--, el resto es pan comido, dicen los más optimistas. En cualquier caso y pase lo que pase con aquel ámbito orético tan poco lucido para la psicología, aquí estamos los hombres hechos unos prodigiosos y potentísimos ordenadores digitales que son la envidia del universo. Sólo eso. Si ésta no es una tesis con un alcance antropológico indiscutible, entonces no hay ninguna que lo sea. En el fondo de esta tesis yace la intuición de que lo que en ella se discute es mucho más que lo que explícitamente aparecerá como un examen de los problemas técnicos insoslayables en la fundamentación de la psicología, *pace* la ironía estilística anterior.

Pensarnos a nosotros mismos como insertados en un mundo circundante compuesto de cosas y de otros sujetos que están ahí; pensarnos, en fin, en un mundo al que pertenecemos como partes suyas y como partes que establecen relaciones causales con él a fin de sobrevivir durante un lapso de tiempo más o menos amplio, aunque siempre limitado, es, seguramente, una creencia tan sumamente básica, que renunciar a ella supondría un auténtico cataclismo intelectual, como decía Fodor. Esta creencia, prototeoría o, quizá incluso, teoría legítima, es la psicología popular. Investigar los fundamentos y supuestos bajo los cuales es posible esta clase de creencias tan sumamente arraigadas, ha de ser parte integral de la labor filosófica.

CAPITULO 2.

PSICOLOGIA, METODO Y MENTE METODOLOGICA.

& 1. Presentación.

El resultado provisional de la discusión del capítulo precedente había sido la conveniencia de decantarse en favor de Sócrates y Fodor, en relativa oposición a Husserl, para reconocer que dos de las propiedades fundamentales con las que se nos presenta el ser psíquico desde el punto de vista del sentido común, del mundo de la vida--en un sentido modificado respecto del concepto fenomenológico estricto--, son su intencionalidad y su capacidad causal. Este suelo descriptivo básico, tomado como necesario fundamento de la inteligibilidad de la psicología científico-objetiva (si es que hay tal cosa), tenía la función de proporcionar la primera circunscripción generalísima de los objetos psicológicos y de sus propiedades. Su misión, pues, era la de abrir provisionalmente el horizonte entero de investigación de la psicología. Este resultó ser el de las actitudes proposicionales, con su peculiar isomorfismo entre relaciones intencionales y causales. Finalmente, el problema que había quedado aleteando en espera de ser abordado más detalladamente fue el de averiguar cuáles serían los auténticos compromisos ontológicos de una ciencia psicológica que se construyera teniendo como punto de referencia orientativo este suelo descriptivo básico insuprimible y que avanzara guiada por ciertos supuestos

interpretativos o esquemas heurísticos que permitieran el descubrimiento progresivo del "mundo natural". Semejante ciencia psicológica habría de dar mejor cuenta de este mundo natural en términos, no ya "populares", sino científico-objetivos. Son estos compromisos los que proporcionan el fundamento ontológico del ser psíquico; lo que el ser psíquico *verdaderamente* es. Este capítulo tiene puestas sus miras en la reconstrucción de estos compromisos ontológicos. Analizarlos es lo mismo que exhibir cuáles serían las propiedades y estructuras de los objetos y estados de cosas hipotéticos a los que se refieren los enunciados sobre los que cuantifica la teoría psicológica.

Es indispensable reconocer, sin embargo, que llegar hasta este análisis requiere una serie de etapas intermedias que están sembradas de trampas. Entre éstas, primordialmente hay dos. La primera de ellas es saber a qué psicología, entre las múltiples clases de investigaciones que históricamente se han llamado así—con una adjetivación u otra—, hay que atender para desvelar sus compromisos ontológicos. ¿Al psicoanálisis acaso? ¿A la psicología fenomenológica de la *Gestalt*? ¿Quizás al conductismo? Y si a este último, ¿a cuál? ¿Al neoconductismo metodológico de Hull y Tolman ¿Al conductismo de Skinner? ¿O entre todas las teorías y pseudoteorías que dicen llamarse "psicología" debemos atender a esa que se conoce hoy día como "psicología cognitiva"? La segunda dificultad radica en la defensa racional de que la psicología elegida entre todas es, por cierto, una ciencia hipotético-deductiva, porque pudiera suceder que ni siquiera lo fuera. En este último caso, nada de lo que se ha dicho en el capítulo anterior llevaría muy lejos. Si ésa fuera la situación final, es claro que no tendría ningún sentido que la filosofía objetivista intentase hacerse con los supuestos implícitos de aquella psicología cuyo punto de partida inteligible fueran las propiedades que atribuimos irremediable e ingenuamente al ser psíquico y su destino fuera la explicación objetiva de éste. Sería un mapa muy bonito

para el que no habría realmente ningún terreno representado que explorar. Sólo una vez que se hayan salvado estos dos escollos es posible dedicarse con cierta confianza a desvelar las características de las entidades con las que está ontológicamente comprometida tal psicología objetiva, cota máxima de este capítulo.

Como puede intuirse fácilmente, adentrarse minuciosamente en cada una de estas etapas requeriría de por sí una monografía independiente. Pero, por fortuna, puede dejarse parcialmente de lado esta empresa enorme. En su lugar, bastará centrarse en algunos hitos seleccionados del desarrollo de estos problemas, por lo demás, tan ampliamente discutidos durante decenios por numerosos filósofos de la ciencia, filósofos de la mente y psicólogos profesionales. Todo este trabajo previo de generaciones permite presentar en estas páginas el extracto sumarisimo de los términos generales en los que se ha planteado la discusión. Es claro, por consiguiente, que en ningún momento esta manera de proceder debe dar la falsa impresión de que se pueden presentar los resultados definitivos de la controversia y que, a partir de ahora, ya no hay nada más que debatir. Ciertamente la sensación que se tiene cuando se lee la bibliografía sobre estos problemas es, justamente, la contraria. Se puede discutir y matizar hasta la última palabra.

Sin embargo, la preocupación más urgente en este instante, mucho más que tomar estas modestas precauciones iniciales, es saber cómo habría que proceder en relación con las dos dificultades planteadas más arriba. La desorientación en el punto de partida mismo puede ser casi absoluta y producir una sensación de cansancio antes siquiera de iniciar la labor prevista. Propongo eludir provisionalmente este inconveniente adoptando la siguiente estrategia: (i) ceder enteramente a las incitaciones de aquella psicología que se nos presente hoy día con el atractivo señuelo de ser la ciencia hipotético-deductiva del ser psíquico, (ii) defender contra las críticas

su carácter efectivo de ciencia objetiva, y (iii) exhibir finalmente sus compromisos ontológicos. Por ser una estrategia, no se excluye de antemano que finalmente todo salga rematadamente mal y que haya que empezar de nuevo. Sea.

Uno de los peldaños de la tarea global que se acaba de anunciar ha de ser necesariamente—y, quizás, desgraciadamente—la espinosísima cuestión acerca de cuál es "el método de la psicología". Tener unas cuantas buenas ideas acerca de este punto decisivo sería sumamente útil para desarrollar la estrategia que propongo. Pero no hay terreno más movedizo que éste, sobre todo si se adopta respecto de él una perspectiva histórica. Es así que, por fuerza, se necesita reconocer una vez más que sería una osadía exagerada pretender analizar este problema en toda su vasta extensión y riqueza de matices. Sin embargo, no me daré por satisfecho si no quedan suficientemente descritos cuáles son algunos de los rasgos genéricos fundamentales con los que dicho método se ha ido cargando, solidificando y depurando poco a poco a lo largo de este siglo, y sin los cuales no sería posible avanzar en el programa presentado.

De acuerdo con la estrategia propuesta, hago primero una serie de afirmaciones en cadena sin justificar. Primera: sucumbo teóricamente por completo a la incitación de la psicología cognitiva y la señalo como la mejor ciencia psicológica con la que contamos. Segunda: el método de la psicología cognitiva es el método hipotético-deductivo de las ciencias objetivas. Lo que es lo mismo: la psicología cognitiva es una ciencia natural-objetiva y, como tal, aplica el método de las ciencias especiales. (Obsérvese que digo "aplica" y no que "debe aplicar" o que "tiene la pretensión de aplicar" el método postulacional si es que quiere erigirse en una ciencia objetiva. Han quedado atrás los tiempos pioneros en los que los psicólogos conductistas pusieron todo su empeño en hacer de la psicología una ciencia natural.) Tercera: los compromisos de la psicología

cognitiva lo son con el funcionalismo y el computacionalismo. Estos compromisos explicarían idealmente las dos propiedades esenciales del ser psíquico a las que se ha aludido constantemente. Mi interés es ver qué sucede cuando se adopta esta estrategia y alcanzar a divisar hasta dónde se puede defender racionalmente.

& 2. Primeros escarceos.

Cuando se señala a la psicología cognitiva como la mejor teoría psicológica que se puede encontrar en el entorno actual de teorías científicas, se quiere aludir, al menos idealmente, a un cuerpo de conocimientos acerca de ciertas entidades y estados de cosas, así como a una estrategia general de investigación racional que permite adquirir estos conocimientos e, incluso, dar un nuevo sentido a muchos otros ya existentes. De acuerdo con la presentación previa, estas estrategias de investigación habrán de estar dirigidas hacia el análisis sistemático de la pregunta que inquiere por el ser psíquico, por su "naturaleza". Se presenta, pues, como un bloque homogéneo de pesquisas racionales, saber qué es verdaderamente el ser psíquico, desvelar cuál es esta estrategia de investigación que permitiría su descubrimiento progresivo, al menos por lo que hace a sus rasgos predominantes, y, finalmente, echar una mirada al cuerpo de conocimientos que define y explica objetivamente el ser psíquico en función de dicha estrategia.

Uno de los hitos fundamentales en la historia de este problema conjugado, un hito que supuso un viraje de gran alcance respecto de la tradición metodológica conductista en psicología, es la obra que Wilfrid Sellars publicó en las décadas de los años cincuenta y sesenta, justamente cuando el conductismo encaraba algunos problemas de fondo muy graves. Es él, por otra parte, uno de los pensadores que más ha influido, según

creo, en la modelación del estilo general que adoptaron a partir de entonces muchas discusiones importantes en filosofía de la psicología. Entre otras contribuciones significativas, este filósofo ha expuesto, de una forma condensada y rigurosa, algunas de las líneas directrices centrales que deben orientar todo método que permita abordar, desde el punto de vista científico-natural, la pregunta que inquiere por "la naturaleza" de los pensamientos así como, en general, por la naturaleza de las representaciones y los procesos mentales¹.

La pregunta "¿qué es un pensamiento?" es, a buen seguro, una de esas típicas preguntas que, según Wittgenstein, produce un espasmo mental agudo en quien la escucha por primera vez y que parece atrofiar con carácter indefinido cualquier esfuerzo sincero que quiera hacerse para darle una respuesta con visos de plausibilidad racional. Simplemente parece que no sabemos siquiera por dónde empezar y que nuestras ideas, aun antes de comenzar, dan vueltas de un lado para otro sin rumbo ni guía fijos. Contraviniendo quizás temerariamente el consejo de Wittgenstein (1984, 27), dejémonos sugestionar siquiera por un momento--metodológicamente, como se suele decir--por la aparente comprensión que se produce en nosotros cuando oímos el sustantivo "pensamiento" y busquemos para él un objeto correspondiente. ¿No queremos *referirnos* a algo cuando utilizamos este sustantivo? Pero, ¿a qué exactamente? ¿Acaso al carácter genérico de todos los modos de la sustancia inextensa de Descartes? ¿A las múltiples especies de *cogitationes* del *Ego* husserliano? Para desbloquear el atrofiamiento primero en que nos ha podido sumir la pregunta así tan directa y planamente formulada, Sellars propone abrazarse a la siguiente estrategema. (Es necesario advertir desde ahora mismo que al tratarse aquí de una

¹ A partir de ahora se entiende que el término "pensamiento", tal y como lo utilizan Sellars y otros muchos autores, es un membrete genérico para referirse al ser psíquico. Los recuerdos, las percepciones, los procesos de aprendizaje, los actos lingüísticos son, en este sentido, pensamientos.

especie de astucia, su viabilidad sólo podrá ponderarse en cuanto a su rigor teórico y utilidad práctica por los resultados que ella dé a la larga en la investigación científica concreta.) Pudiera suceder que no sepamos decir nada a bote pronto cuando se nos pregunta qué es un pensamiento, excepto que los tenemos-es decir, que pensamos esto o aquello-, y que sabemos que los tenemos, es decir, que sabemos que pensamos esto o aquello. Este saber que los tenemos, dicho sea de paso, tiene la extraordinaria peculiaridad de que nos lleva a concluir que hay pensamientos sin necesidad alguna de argumentaciones posteriores para establecer la verdad apodíctica de esta afirmación de existencia. En el giro cartesiano-husserliano, esta idea se expresa diciendo que tenemos la certeza apodíctica de que estamos pensando esto o aquello en el momento mismo en que lo estamos pensando. El aturdimiento primero parece que se ha aliviado un poco con estas afirmaciones. A lo mejor no sé lo que son los pensamientos, pero desde luego sé que los tengo. Yo, cuando pienso esto o aquello, sé que lo estoy pensando y puedo, además, expresarlo con palabras. Yo sé que *ego cogito cogitata*. Sin embargo, este momentáneo alivio no debe darnos la falsa sensación de haber alcanzado un triunfo teórico especialmente significativo en relación a la pregunta planteada. Que yo sepa que pienso y que tengo pensamientos con una nitidez epistemológica única, no quiere decir que haya contestado a la pregunta: ¿qué es un pensamiento? Como mucho he contestado a las preguntas: ¿hay pensamientos?, ¿cuáles son sus rasgos descriptivos fenomenológicos generales?, ¿cómo sé que hay pensamientos?, ¿cuál es la transparencia epistemológica de este conocimiento? Ciertamente, no es en absoluto un asunto baladí responder con rigor a estas cuestiones, especialmente desde la perspectiva de quienes estén directamente interesados en ciertas formas de la teoría del conocimiento. ¿Quién lo negaría? Aun así, no tengo derecho a albergar ninguna seguridad respecto de la posibilidad de saber algo positivo acerca de lo que habría que decir como

réplica a la interrogación "¿qué es el pensamiento?" por mucho que tuviese a mi disposición una teoría rica y bien articulada acerca de las preguntas antes formuladas. Rorty, por ejemplo, quien ha reconocido su deuda intelectual con Sellars, piensa ciertamente así (Rorty 1982a, 331).

Antes de continuar quiero repetir una vez más que la tentación de plantear la pregunta acerca de la naturaleza de los pensamientos pudiera responder simplemente a una debilidad crónica, cuyo origen no sería otro que el de ceder al intento de investigar las ilusiones típicas que, curiosamente, sólo se erigen inadvertidamente ante nuestros ojos cuando se formulan "cuestiones filosóficas". Pero de ilusiones también se vive, dice el refrán. Y a Sellars, Fodor y Chomsky, entre otros muchos, no parece que les importe demasiado vivir de ilusiones, al menos hasta que no se demuestre fehacientemente que lo son; cuestión ardua donde las haya.

Con el fin de ir poniendo de relieve poco a poco cuál es el giro exacto que la filosofía objetivista ha adoptado en los últimos decenios en la investigación del ser psíquico, me interesa destacar que la diferencia de fondo más importante entre ella y la filosofía de la subjetividad de la tradición cartesiano-husserliana es la siguiente: preguntar *qué* es el ser psíquico es una pregunta legítima y a ella sólo puede contestar la estratagema y astucia bien fundamentada que consiste en no dar una importancia especial a las cuestiones relativas al carácter fenomenológico y subjetivo de las vivencias intencionales y a los grados de evidencia con el que se nos dan las proposiciones relativas a ellas y a sus objetos. Consecuencia inmediata de esta diferencia es la posibilidad de poner en tela de juicio la "autoridad epistemológica final" de la primera persona (Rorty 1970, 208). Sirvan estas dos indicaciones de anuncio programático para investigaciones posteriores bastante más complicadas e interesantes, ya que en ellas se tratarán cuestiones de principio relativas a la psicología. De

momento, es evidente que hay que avanzar mucho más todavía para ver con claridad la maraña de problemas aquí involucrados.

Volvamos a retomar con renovado ahínco, en el más puro estilo socrático, la tarea planteada. ¿Qué son los pensamientos y qué propiedades tienen? Tú dices que los tienes y ahora yo te pregunto qué es eso que tú dices poseer con absoluta seguridad. Que siga siendo Sellars nuestro interlocutor en este diálogo. Este filósofo considera que, en una segunda reflexión, hay dos notas que se adelantan y que permiten comenzar tímidamente la investigación propuesta. Tienen, además, la peculiaridad de que admitirían ir más allá de las reflexiones de cuño cartesiano antes esbozadas. La primera de esas notas es que los enunciados de los lenguajes naturales son la *expresión* pública de los pensamientos. La segunda, consecuencia indirecta de esta primera, es que las propiedades de los pensamientos son *análogas* a las propiedades de los enunciados del discurso público. Tomemos estas dos propiedades como puntos de apoyo y veámos hasta dónde nos conducen. Por lo demás, estas dos notas que Sellars señala se darán con el carácter de *déjà vu* a quien conozca las tesis de la tradición filosófica acerca de las relaciones entre el lenguaje y el pensamiento. Nada esencialmente nuevo está presente aquí, pero estas dos notas han sido abrazadas con un renovado entusiasmo por sectores muy influyentes en la filosofía de la mente contemporánea.

Después de señalar estas dos notas, Sellars aconseja a continuación dejar de lado momentáneamente el hecho mencionado más arriba de que sabemos que tenemos pensamientos; esto es, que los pensamientos, como él dice, se nos dan *aparentemente* en "cualidad" autoconsciente. En su lugar, propone que nos centremos sobre todo en que los concebimos, una vez más, como "sucesos internos que son análogos al habla (speech)" y que encuentran su expresión pública en el habla--aunque, por supuesto, puedan suceder en ausencia de esta expresión pública" (1963, 31-32). En un

principio, es más manejable y se mostrará teóricamente más fructífero, al menos por lo que hace a los fines de destacar las estrategias de investigación generales de la psicología cognitiva, la consideración de los pensamientos de esta segunda manera, que hacerlo de la primera—como algo que se nos da *aparentemente* en "cualidad" consciente. Digo siempre que los pensamientos se nos dan aparentemente de esta manera, porque para Sellars es "un error suponer que la autoconciencia del pensamiento conceptual se nos presente a sí misma de una manera cualitativa" (1963, 32) a diferencia de lo que pasa, al parecer, con las sensaciones, las imágenes, los dolores, los picores, etcétera (Ibid; 1972a, 198)². Sin embargo, es curioso, y merece la pena constatarlo desde el inicio, que Sellars defienda, en unión con esta caracterización de la "naturaleza" de los pensamientos como algo que no se nos da realmente de una manera cualitativamente consciente, el carácter no inferencial, no deductivo, del conocimiento que dice que los pensamientos son análogos a los enunciados y que quedan, en numerosas ocasiones, legítimamente expresados en ellos.

...el conocimiento directo que tenemos de lo que estamos pensando <cuando pensamos que hace frío fuera> es que algo análogo y propiamente expresado en el enunciado "hace frío fuera" está sucediendo en mí. (1963, 33).

Con el carácter autoconsciente de los pensamientos no sabemos por ahora qué hacer, excepto constatar la presuposición de que con la expresión del pensamiento, tanto a mí mismo como a los demás, también lo expresamos. El "se dice" cuando expreso el pensamiento "hace frío fuera". Por otro lado, si en una situación de comunicación normal alguien nos

² Sin duda alguna, estas ideas hubieran merecido una aclaración mucho mayor por parte de Sellars para que fueran completamente inteligibles. No conozco, sin embargo, ningún lugar en su obra filosófica al que poder acudir para arrojar definitivamente luz sobre esta cuestión.

dice que "hace frío fuera", inmediatamente le atribuimos la autoconciencia: él sabe o cree que hace frío fuera. Por contrapartida, explotar el carácter, dado al parecer inmediatamente, de la analogía entre el lenguaje y el pensamiento conduce a resultados que han dado un juego enorme en la metodología psicológica cognitiva. Es esencial retener y fijar este giro definitivo que han adoptado las investigaciones acerca de los pensamientos si es que se quiere comprender adecuadamente cuáles son los problemas de fundamentación radicales que aquejan a la psicología cognitiva. Dentro de este giro está la consideración de la *naturaleza* de los pensamientos como algo a lo que no se tiene ningún acceso directo inmediato--al contrario, quizás, de lo que pasa con su *existencia*--, sino sólo analógico (Sellars 1963, 33; 1972b, 323), así como que ellos se expresan en muchas ocasiones con los enunciados de los lenguajes naturales que utilizamos y comprendemos. La naturaleza de los pensamientos es algo que en parte se muestra en su expresión comprendida, pero que en parte se oculta en su naturaleza completa.

La analogía entre el lenguaje y el pensamiento se ilumina metafóricamente mediante otra nueva. Sellars propone que pensemos en la siguiente situación. Sin movernos de casa, podemos seguir fielmente los movimientos de una partida emocionante de ajedrez que está celebrándose en un país extranjero. Para que este seguimiento sea fiel--para que sea un seguimiento de *esta* partida de ajedrez--, basta con que se dé la siguiente condición. Que nuestro concepto de las piezas y del tablero de ajedrez, así como también el concepto de los movimientos reglamentados de las piezas sobre el tablero en el país extranjero, *cumplan un papel análogo* al de las piezas, tableros y movimientos que tienen lugar cuando *nosotros* jugamos al ajedrez. Esta es la condición esencial, y su formulación concreta tiene su razón de ser en que sabemos que las piezas del ajedrez pueden tener diversas formas, colores y tamaños, y que los tableros de ajedrez pueden

estar colocados en múltiples posiciones, o incluso muy deformados, con tal de que preserven ciertas cualidades topológicas generales. De la emocionante partida, a nosotros sólo nos llegan, por ejemplo, las siglas P-K3 que, de acuerdo con la notación convencional inglesa y las reglas del ajedrez, es la expresión del movimiento de cierto peón de una posición determinada a otra. Pero de las propiedades reales, intrínsecas, de las piezas que han movido los jugadores sobre el tablero en el país extranjero, así como de los movimientos físicos concretos de estas piezas sobre su superficie, no tenemos idea exacta alguna.

Con los pensamientos, dice Sellars, ocurriría tres cuartos de lo mismo.

Nuestro concepto de "lo que son los pensamientos", al igual que nuestro concepto de lo que es un enroque en ajedrez, es un concepto abstracto en el sentido de que no concierne en sí mismo al carácter *intrínseco* de los pensamientos, *excepto como objetos que pueden suceder en patrones de relaciones que son análogos a la forma en que los enunciados están relacionados unos con otros y con los contextos en los que se usan* (1963: 34. Véase también Armstrong 1977, 89).

Del carácter intrínseco de los pensamientos y de sus interesantísimos movimientos y relaciones sólo nos llegan, para apurar un poco más la metáfora, ciertos signos—los enunciados de los lenguajes naturales—, que son la expresión "de los movimientos" relativos de los pensamientos. Ha quedado así inaugurada para el uso y discusión de las generaciones venideras la tesis del "papel conceptual" (*conceptual role*), tanto en filosofía de la mente como en semántica³. Dado que las propiedades esenciales de los enunciados de los lenguajes naturales son las de estar sintácticamente

³ Una explicación más adecuada que dé cuenta del carácter exacto y de las funciones de los papeles conceptuales se encuentra en el & 3 (b) de este capítulo.

bien formados y poseer un contenido semántico, así también los pensamientos y, en general, las representaciones mentales, en tanto que objetos análogos a los objetos lingüísticos por excelencia, habrán de tener una sintaxis y una semántica o, si se quiere expresar así, sus necesarios análogos. (La psicología cognitiva, desde el momento en que adopta este formato explicativo general de carácter lingüístico, es una teoría analógica y abstracta en el nivel de análisis en el que establece sus explicaciones. Como se indicará inmediatamente y se justificará posteriormente, su índole abstracta explicativa lo es en relación al carácter "concreto" ontológico de ciertos mecanismos y procesos neurofisiológicos.) El contenido semántico y la forma sintáctica de las representaciones mentales estarán determinados por el papel que cumplan los enunciados y sus significados—las proposiciones—en ciertos procesos cognitivos típicos tales como, por ejemplo, las inferencias. En la terminología de la filosofía de la ciencia, los enunciados y sus significados funcionarían para la psicología cognitiva como *modelos* con los que poder empezar a explicar de una manera abstracta, pero relevante racionalmente, las propiedades de ciertas entidades no observables directamente: las representaciones mentales.

Un ejemplo del alcance y empleo de esta estrategia abstracta y analógica es el programa de investigación que propone Chomsky para la lingüística y la psicología contemporáneas, campos de investigación que, según él, no son en absoluto independientes. Sólo dos pinceladas. En primer lugar, Chomsky quiere que seamos sensibles al hecho de que en el estudio del lenguaje "procedemos abstractamente" (1988, 8). Los conceptos más adecuados para la explicación y predicción en lingüística seguramente seguirán siendo los tradicionales conceptos abstractos de oración, nombre, verbo, complemento, etcétera, por mucho que avance nuestro conocimiento de las propiedades y estructuras cerebrales concretas que intervienen, supuestamente, de una manera necesaria en los procesos lingüísticos. En este

sentido, el estudio del lenguaje y de la mente es el estudio de las propiedades abstractas de mecanismos cerebrales desconocidos (Chomsky 1988, 8). En segundo lugar, la forma de acceder al estudio del lenguaje y de la mente, y en particular, al estudio de la facultad lingüística como una facultad particular de la mente, es partiendo siempre del análisis y comparación de las oraciones que los hablantes de un idioma consideran estructuralmente correctas y semánticamente significativas. Desde esta base empírica, el lingüista avanza un paso más y construye la gramática de ese lenguaje particular para, finalmente, llegar a alcanzar algún conocimiento de los principios formales absolutamente generales que constituyen la gramática universal, esto es, el estado de la facultad lingüística antes de toda experiencia de los hablantes con un lenguaje real (Chomsky 1988, 60-61; para más detalles véase entero el capítulo segundo de esta obra de Chomsky). Con estas dos indicaciones a la vista, se puede colegir la importancia y alcance efectivos que la estrategia que esbozó Sellars sigue teniendo hoy en día para diseñar programas de investigación de gran alcance tanto en psicología como en lingüística. Por lo demás, lo que le interesa al filósofo es, precisamente, fijar estas cuestiones metodológicas de principio y discutir las crítica y radicalmente. En cualquier caso, sería imposible descender al detalle de las teorías actualmente en vigor que tratan sobre estas cuestiones y cuyo objeto, tal y como se ven ellas a sí mismas, es el estudio de la mente y de sus diversas facultades. Simplemente, mi interés es el de subrayar suficientemente el conjunto de metáforas y analogías operantes que dirigen la elaboración de las diversas teorías.

Aunque estemos dispuestos a aceptar provisionalmente, en razón de su verosimilitud, la estrategia que propone Sellars, falta todavía saber qué se puede hacer con ella a partir de este momento para que sea de alguna utilidad teórica y empírica en psicología. ¿Cómo explotar el carácter analógico entre el lenguaje y el pensamiento para que esa analogía rinda

frutos teóricos objetivos válidos? Desde luego, este rendimiento deseado sólo se puede alcanzar si se logra armonizar y fundir la analogía dentro del marco explicativo ideal que ofrece la metodología hipotético-deductiva de las ciencias objetivas. Naturalmente, buscarle un hueco adecuado a esta analogía dentro del método postulacional requiere como paso previo decir unas palabras que sinteticen cuál es este método. Sólo así no quedarán todas estas reflexiones demasiado en el aire y se podrán discutir con más detalle.

& 3. El método de la ciencia natural y el método de la psicología cognitiva.

(a). El método de la ciencia natural: rasgos genéricos.

La distinción de la que normalmente parte la filosofía de la ciencia para hablar del método científico, es la que se da entre el lenguaje teórico y el lenguaje de observación. Dentro del lenguaje teórico se formulan los principios y axiomas fundamentales de la teoría que quiere construirse. A él también pertenecen los enunciados que se deducen a partir de este grupo de principios y axiomas máximamente universales. La característica primordial de estos enunciados es que ellos se introducen como enunciados primitivos por los principios y axiomas de la teoría en relación con los enunciados de observación. Estos enunciados teóricos se refieren hipotéticamente a un dominio determinado de entidades postuladas cuyas propiedades y estructuras básicas quedan establecidas también por los principios fundamentales de la teoría. La teoría sólo quedaría idealmente completada en su construcción--la teoría es sólo teoría--si se consigue correlacionar, bajo ciertas condiciones muy complejas, estas entidades teóricas con ciertos objetos y estados de cosas directamente observables.

A estos estados de cosas no se referirían directamente los enunciados de la teoría, sino los enunciados observacionales.

La interpretación exacta que merecen estas correlaciones entre las entidades teóricas y las entidades observables es, por cierto, una de las cuestiones más perturbadoras dentro de la discusión metacientífica y en donde el teórico de la ciencia puede ser más escrupuloso. La identificación simple entre las entidades postuladas por la teoría y los estados de cosas observables--la temperatura es la energía cinética media de las moléculas--la redefinición operacional del lenguaje teórico en el lenguaje observacional y, como continuación de ésta, el análisis lógico de los enunciados de la teoría en términos de enunciados observacionales, son posibilidades que quedan abiertas cuando simplemente se habla de correlaciones. Qué opción sea finalmente elegida determinará una corriente metacientífica u otra. Sin embargo, sea lo que sea de esta intrincada cuestión hay otras dos relacionadas con ella que no se pueden dejar de lado. Me refiero, en primer lugar, al hecho de que a los estados de cosas postulados se les dota racionalmente con la función de ser las causas de los estados de cosas observados que se toman como efectos y "signos" de aquéllos. Los electrones, de acuerdo con la teoría física disponible, son las causas de los trazos que se observan en la cámara de niebla. En segundo lugar, importa destacar siempre que las correlaciones entre los estados de cosas hipotéticos y los estados de cosas observables no son ellas mismas observables, sino también postuladas (Chihara y Fodor 1991, 146). La justificación de estas correlaciones postuladas dependería fundamentalmente de criterios epistemológicos generales como los de simplicidad, plausibilidad y adecuación predictiva de los sistemas explicativos tomados como un todo, mucho más que de criterios firmes que se puedan aplicar rígidamente en todas las situaciones (Chihara y Fodor 1991, 140).

A esta exposición primera hay que añadir todavía que la correlación postulada no se puede efectuar sin más ni más entre los enunciados observacionales y sus referentes, y los enunciados teóricos y los suyos. Los enunciados observacionales particulares sufren una primera elaboración antes de ser puestos en correlación con los enunciados teóricos. Esta elaboración consiste en buscar una nueva clase de correlaciones funcionales empíricas entre los referentes de los enunciados observacionales. Estas correlaciones se expresan mediante enunciados condicionales de la forma bien conocida: si en un tiempo t y en ciertas condiciones iniciales, se dieran ciertos estados de cosas, entonces se observarían tales y cuales otros estados de cosas, *ceteris paribus*. Estas correlaciones permiten formular, sin salir todavía del lenguaje observacional, leyes inductivas. Son estas leyes inductivas, que expresan generalizaciones empíricas, y no enunciados observacionales sueltos, las que se correlacionarían finalmente con ciertos grupos suficientemente amplios de proposiciones y teoremas que se deducen de los principios y leyes universales de la teoría. (Quine afirma que realmente es la teoría entera, en tanto que sistema holístico, la que se somete a contrastación empírica.) Una teoría científica es explicativa cuando tiene la capacidad de poder "derivar" enunciados que se consideran las contrapartes teóricas de las leyes inductivas a partir de un grupo reducido de principios y leyes teóricos.

Dentro de esta presentación breve, idealizada y meramente "lógica" del método hipotético-deductivo, no conviene dejar de lado otro matiz importante que completaría y haría más real el funcionamiento esquemático del método científico, y la comprensión de la formación de teorías científicas. Consiste éste en señalar la importancia que tienen los *modelos* para el método de la ciencia y la necesidad de encontrar uno adecuado para poder empezar a construir y formular los principios de la teoría por algún sitio razonable e inteligible. Los principios de la teoría no se

introducen en ningún caso con la intención de que cumplan el papel de ser meros formalismos enteramente asignificativos con los que poder jugar a hacer cálculos racionales; cálculos que, en el mejor de los casos, se podrían poner posteriormente en correlación con las leyes inductivas de generalidad inferior. Respecto de los principios de la teoría, los modelos tienen, antes bien, la función de conducirnos orientativamente de lo conocido a lo desconocido. Es por este motivo por el que los modelos son instrumentos de un valor inestimable al servicio de la elaboración de las teorías científicas. Un modelo es, en este sentido, simplemente una descripción articulada de un dominio de objetos de los que se conocen sus propiedades más relevantes. Los modelos permiten entender cómo serían y en qué consistirían los fenómenos que tienen que ser explicados si consistieran en la clase de cosas de las que ellos hablan (Sellars 1972a, 204). Sin embargo, no debe confundirse en ningún momento el modelo que se adopta como patrón conocido con el sistema de los supuestos teóricos que constituyen los principios y leyes más universales de la teoría. Los modelos se anuncian siempre con una carta de presentación que les resta competencias y que dice así: la analogía entre las propiedades de los objetos conocidos de los que nosotros hablamos y las propiedades de las entidades postuladas por la teoría a partir de sus principios y leyes fundamentales, es siempre limitada y está necesariamente matizada.

(b) **Modelos lingüísticos.**

Cuando se tienen presentes estos elementos mínimos del método postulacional, es cuando se puede intentar insertar la estrategia que Sellars propugna para la psicología dentro del marco explicativo científico general. Ya se ha mencionado que el modelo, o uno de los modelos principales de la teoría psicológica contemporánea, son los enunciados de los lenguajes

naturales y sus propiedades y relaciones sintácticas y semánticas. En particular, las proposiciones, clásicamente consideradas como los contenidos semánticos de los enunciados, tienen para la teoría psicológica la función de proporcionar *índices externos* (Schiffer 1989, 25) que permitirán adscribir papeles conceptuales abstractos a los "pensamientos". En este sentido, en cuanto índices externos, la función de las proposiciones respecto de los estados causalmente implicados en la etiología del comportamiento es análoga a la función que tienen los números respecto de las propiedades físicas de los cuerpos. También los números son índices externos para la elaboración de las teorías físicas. La analogía ha sido claramente expuesta por Stalnaker. Si es correcto, como parece que lo es al menos desde Galileo, representar y entender las propiedades físicas que se adscriben a los objetos físicos como relaciones entre éstos y los números, es porque las propiedades físicas tienen "una estructura en común" con las propiedades de los números (Stalnaker 1984, 9). Sencillamente, usamos un número determinado para representar un peso o una altura determinada en tanto que propiedades de este o aquel objeto físico. Así, de la misma manera que las propiedades familiares de los números sirven de modelo para elaborar la teoría física de la materia, las propiedades familiares de los enunciados sirven para elaborar la teoría psicológica de la conducta.

La ventaja de los modelos numéricos es que ellos han mostrado ser históricamente insuprimibles y de una eficacia innegable para la teoría física. En el caso de la psicología, hay que esperar todavía para lograr ver hasta dónde llega el alcance y la eficacia empírica de los modelos lingüísticos, aunque se pueda decir ya que han dado frutos innegables. Así, por mencionar sólo un ejemplo, los modelos de Rumelhart y McClelland como modelos psicológicos del razonamiento inductivo. Pero en este momento, antes que de hacer las alabanzas correspondientes a la investigación ya en marcha y a sus resultados, lo único que se trata de afirmar es que no hay

ninguna razón para no aplicar los modelos lingüísticos a los procesos psicológicos y que si hay que desestimar esta aplicación será porque sea inútil para elaborar teorías psicológicas explicativas, y no porque haya en contra suya razones *a priori* de peso. De esta manera indirecta es como entrarían los modelos lingüísticos analógicos dentro de la explicación psicológica.

Lo que tenemos hasta este momento es lo siguiente. Preguntar por el ser psíquico desde el punto de vista de la filosofía objetivista que se apoya en el método postulacional de la ciencia conduce a la explotación de modelos lingüísticos en psicología. Estos modelos lingüísticos proporcionan una herramienta heurística para la investigación sistemática de las propiedades de los estados de cosas postulados por la ciencia psicológica como causas de la conducta. El ser psíquico—el isomorfismo entre causalidad e intencionalidad—, si es que se deja investigar objetivamente, será porque se pueda captar su naturaleza cuando se siguen las consecuencias de las líneas directrices que aquí comienzan a despuntar.

Sin embargo, a pesar de estas primeras consideraciones, es muy probable que todavía no se vea de una forma suficientemente diáfana cómo es que este modelo hipotético-deductivo llega a funcionar para el caso de la psicología. Hasta ahora, se ha hablado en general del método hipotético-deductivo y de la función que los modelos cumplen en él, así como de la posibilidad abierta de aplicar modelos lingüísticos en psicología. Falta en este momento, sobre todo, hablar con más detalle del lenguaje y de los hechos de observación que están a la base de la construcción psicológica científica.

Para rondar un poco más cerca del análisis que sería necesario emprender aquí, es bueno recoger parte de lo que se ha sembrado. El fundamento de la inteligibilidad de la tarea psicológica científica a propósito de los estados y procesos mentales serán siempre los recursos conceptuales

más o menos refinados que ofrezca la reflexión sobre "la" psicología popular. Sobre este suelo preteórico, se alzaría la construcción hipotética de la ciencia. Pero, ¿de qué manera exactamente? Son principalmente dos las cosas a tener en cuenta aquí. La primera se refiere a las restricciones metodológicas que la ciencia impone al sentido común. La segunda, a las correlaciones funcionales empíricas entre los estados de cosas observables que sirven para establecer las leyes inductivas. Decir algunas palabras sobre cada uno de estos dos puntos es prudente.

La restricción fundamental, aunque se podrían enumerar otras más, que la psicología científica impone a nuestros conocimientos psicológicos tácitos no es, en principio, una restricción categorial—acepta una amplia gama de conceptos que pertenecen a y tienen su origen en el sentido común, tales como, por ejemplo, memoria, atención, percepción, aprendizaje, pensamiento, lenguaje, etc.—, sino exclusivamente metodológica. La psicología científica exige que la afirmación de la existencia de estados mentales subjetivos en los organismos cuya conducta investiga el psicólogo se realice sobre fundamentos conductuales. Brevemente y simplificando al máximo, de una persona o de una rata que se chocan sistemáticamente con todos los objetos que, como obstáculos, se hallan en su camino, diremos que no perciben visualmente de una manera adecuada, si es que perciben en absoluto (excluidas, claro está, otras causas e impedimentos). Al menos diremos esta clase de cosas si es que queremos dar una interpretación racional de su conducta. Esta interpretación surge casi espontáneamente y en ella consiste adoptar lo que Dennett ha llamado la "instancia intencional" en psicología: atribuir estados mentales a un organismo es útil para el psicólogo que quiere explicar racionalmente su conducta. También es lo más adecuado. El comportamiento torpe de la persona o de la rata se explica causalmente de una forma sencilla cuando se habla de la ausencia

de estados mentales perceptivos. La instancia intencional es el nivel explicativo inicial adecuado que hay que adoptar para explicar la conducta.

Esta restricción metodológica hace que el comportamiento observado sea tomado sistemáticamente como índice y expresión que señala la existencia o no existencia de los estados mentales que, precisamente, consideramos están implicados en la etiología del comportamiento del organismo. En el ejemplo anterior, el comportamiento que el psicólogo observa en el organismo se toma como evidencia de la no existencia de estados mentales perceptivos. Resulta una afirmación clásica decir que la psicología sólo hace un uso heurístico de la conducta observada del otro, así como, en general, de cualquier otra clase de índice conductual en el que el sistema de las intencionalidades del organismo se haga más o menos transparente.

Por lo demás, los comportamientos observados pueden ser de una gama amplísima. Desde enunciados que expresan los estados que llamamos "subjetivos", hasta movimientos de ojos imperceptibles para el sujeto, pasando por pulsar uno de dos botones diferentes, o el "comportamiento" de la actividad neuronal. Todos estos comportamientos observados se dividen, básicamente, en dos clases fundamentales: verbales (sólo cuando el sujeto sea una persona) y no verbales. En general, por ejemplo, un enunciado introspectivo se tratará como evidencia en favor de lo que una persona está percibiendo, pensando o recordando. Una vez más, la psicología científica está en una relación de continuidad con "la" psicología popular, ya que ambas se apoyan en la evidencia preexistente no cuestionada de que hay correlaciones entre el comportamiento público observable de un sujeto y los estados y procesos mentales que son las causas de algunos de sus comportamientos.

Antes de continuar, importa detenerse un momento para decir unas palabras más acerca de la noción de conducta que utiliza implícitamente la

psicología cognitiva actual. Para ésta, comportarse no quiere decir tan sólo "movimiento" físico, muscular y corporal del organismo, sino "respuesta" en un sentido muy, muy amplio cuyos límites son difíciles de rodear con precisión. Una vez más, la psicología objetiva está en una cierta continuidad con la psicología popular; en este caso, respecto de la noción de respuesta conductual. Por poner un par de ejemplos manifiestos, buscar la solución al problema de la Torre de Hanoi o a un problema criptoaritmético típico cuenta para la psicología cognitiva como "respuestas" del organismo humano ante tareas propuestas.

Correlativamente, qué estemos dispuestos a considerar como respuesta dependerá en gran parte de lo que estemos dispuestos a considerar como estímulo. Si la solución al problema de la Torre de Hanoi o a uno criptoaritmético son respuestas que implican procesos y estados mentales cuyo estudio es relevante para la teoría psicológica cognitiva, entonces tienen que valer como "estímulo" cosas tales como la comprensión de oraciones del estilo "¡soluciona este problema que te propongo!", y la comprensión de la situación problemática misma. A su vez, qué estemos dispuestos a considerar como estímulos y respuestas determinará en gran medida nuestra teoría psicológica y qué fines perseguimos a la hora de hacer ciencia psicológica. Es de la máxima transcendencia tener en cuenta esta afirmación, si no se quiere enredar la discusión con cuestiones acerca de la legitimidad que tiene la psicología cognitiva para hacer lo que hace.

Con estas reflexiones a la vista, es muy probable que, respecto del conductismo metodológico de las décadas precedentes, se haya dado un cambio conceptual en el uso que la psicología cognitiva hace de términos centrales de la teoría psicológica como son los de "conducta observable", "estímulo", "respuesta", "contingencia de reforzamiento" o "entorno experimental" (Hamlyn 1990, 15). Los "estímulos" no son sólo formas de energía a las que son sensibles las terminaciones nerviosas de los órganos

sensoriales del organismo--si es que hay que incluir en absoluto semejante noción de estímulo entre los fenómenos que son psicológicamente relevantes. Tampoco habrá que hacer uso meramente de la noción de estímulo como si fuera cualquier cosa en el entorno experimental que produce una respuesta en el organismo. Antes bien, es preciso incluir esencialmente cosas tales como "aquellas propiedades del ambiente que el animal o la persona percibe de una determinada manera" (Hamlyn 1990, 13). En este último sentido, la tarea de resolver un problema criptoaritmético o el problema de la Torre de Hanoi puede contar como un estímulo para una persona. Para el caso de las respuestas, claro está, se impone la amplificación correspondiente respecto de qué ha de entenderse por tal cosa.

Por lo que se refiere a la aclaración del segundo punto que se necesita para comprender mínimamente cómo surge la construcción psicológica científica del ámbito de la psicología de sentido común, quiero recurrir una vez más a Sellars y a su capacidad para sintetizar largas y arduas cuestiones relacionadas con el método de la psicología "conductista". Así es como él llama a lo que ha venido a ser el método de la psicología "cognitiva" tras varios refinamientos y amplificaciones. Como se acaba de sugerir, la más importante entre estas últimas sería lo que haya de contar como una respuesta y lo que haya de contar como un estímulo. (¿Son los psicólogos cognitivos, aunque a muchos les costaría un disgusto reconocerlo, "la última generación" de psicólogos conductistas?)

Sellars llama propiedades condicionales del organismo ("iffy" properties) a aquellas propiedades que se enuncian genéricamente de la siguiente manera: si en un tiempo t y en ciertas condiciones iniciales ocurriese cierto estímulo, se daría cierta respuesta. Lo que resulta esencial para el método postulacional, cuando éste se aplica a la psicología, es "suponer que las propiedades condicionales del organismo están conectadas

con estados de un sistema de entidades postuladas de acuerdo con ciertos principios postulacionales" (Sellars 1963, 24). Estas entidades postuladas tendrían un estatuto semejante al de las variables intervinientes de Tolman o, sencillamente, al que tienen en general las variables hipotéticas en la ciencia. Ahora se considera que la tarea de la psicología cognitiva ("conductista") es, sin más, la de caracterizar adecuadamente las propiedades de este sistema de entidades postuladas, las cuales formarían parte esencial de los procesos etiológicos que dan lugar a la conducta intencional, inteligente y observable de los organismos. Para empezar a caracterizar estas entidades inobservables, Sellars añade, tal y como era de esperar, que "nuestro conocimiento de transfondo (background knowledge) hace razonable la suposición de que estos hechos condicionales ("iffy" facts) se dan porque está sucediendo un proceso interno que es, en aspectos importantes, análogo al comportamiento verbal público, cada etapa del cual encontraría una expresión natural en el habla pública" (Sellars 1963, 24). Según estas observaciones, es fácil llegar a la conclusión de que el uso heurístico de la conducta de los organismos—en un sentido amplísimo—y la posibilidad de hablar de hechos condicionales del organismo, son cara y cruz de la misma moneda. Pero esta moneda no tiene ningún valor si se desprende de la necesidad intelectual—teóricamente satisfecha por el método postulacional—de encontrar estados de cosas no observables que expliquen causalmente las conductas observadas. Y, en un principio, dada la continuidad categorial entre la ciencia psicológica y el sentido común, mantenemos la creencia de que estos estados de cosas no observables son estados y procesos mentales algunas de cuyas propiedades son análogas a las de los lenguajes naturales.

Las consecuencias que tiene para la psicología el hacer este uso meramente heurístico de la conducta observable no son ni mayores ni menores que las que tiene para otras ciencias el hacer un uso heurístico de ciertos estados de cosas observables. Para verlo con claridad, la mejor

manera de proceder es poniendo un ejemplo extraído de la ciencia física. También en ella hay ciertos hechos observables que se toman como índices de otros hechos no observados, pero postulados para explicar causalmente los observables. Así, por ejemplo, ciertos fenómenos accesibles a la observación inmediata se toman como índices de la presencia de electrones en el entorno experimental próximo. Ciertos trazos observables en la cámara de niebla son un índice, de acuerdo con la teoría física, de la presencia de estos electrones. La matización "de acuerdo con la teoría física" significa que las correlaciones entre los estados de cosas observables—los trazos en la cámara de niebla—y los estados de cosas hipotéticos—los electrones—, son ellas mismas no observables en el sentido que se expuso al comienzo de este párrafo. Sin embargo, están de sobra justificadas si es que se quiere salvar la coherencia, simplicidad y capacidad predictiva de la teoría física como un todo. Ahora bien, si los trazos son un índice de la presencia de electrones es porque los electrones son su causa. Pudiera ser, desde luego, que la teoría física estuviera equivocada y que los electrones no fueran la causa de las manchas características que aparecen en la cámara de niebla y que se interpretan, desde la teoría, como señales de la existencia de electrones en el entorno espacial cercano. En tal caso, estas manchas no deberían contar ya más—no deberían ser interpretadas a partir de entonces—como índices de su presencia. Si alguien, con todo, mantuviera que las manchas en la cámara de niebla siguen contando como índices, nada se le podría objetar excepto que no sabríamos de qué lo son exactamente. Hasta que no contemos con una nueva teoría que nos lo diga; es decir, hasta que la teoría no nos diga cuál es su causa, guardaremos un discreto silencio sobre este asunto. Si no hay teoría, entonces cualquier objeto inmediatamente presente que se tome como señal o índice respecto de la existencia de otro objeto se tomará como tal sin fundamento suficiente. Cuando contemos con una teoría, tendremos realmente índices,

porque la teoría nos dirá cuáles son sus causas. Si no hay teoría, no hay conocimiento de las causas. Si no hay conocimiento de las causas, no hay posibilidad de interpretación fundamentada de un objeto como índice. Por consiguiente, para que un objeto o estado de cosas sea un índice de otro objeto o estado de cosas no es suficiente con que alguien lo "aprehenda" como índice suyo. Si la aprehensión es arbitraria o falsa, es decir, si no hay ningún sendero causal desde el objeto que es señalado al objeto que funciona o se aprehende como señal, este último objeto no sería señal de aquél, aunque así se aprehendiera. La posibilidad de hablar de aprehensiones verdaderas o falsas depende de la existencia o no existencia de senderos causales relevantes. Ningún objeto presente es índice de suyo si no hay causas de su existencia. En otras palabras, una "aprehensión" es parte de una teoría explicativa y no un acto particular aislado o una parte de éste. Por su lado, es la relación causal la que es verdaderamente condición necesaria de la función sígnica, y el conocimiento de esta relación causal es el que nos permite saber qué objetos son índices. La existencia de fósiles, para poner un ejemplo de Husserl, es índice de la existencia de animales antediluvianos, sólo si la existencia y propiedades de estos últimos es la causa de la existencia y propiedades de los fósiles.

Con este ejemplo sólo he querido recordar el hecho trivial de que el procedimiento de tomar un estado de cosas observable como índice y expresión de otro no observable, es connatural a la investigación científica y a su ideal explicativo causal. El objetivo teórico de la ciencia es caracterizar suficientemente los estados de cosas no observables y hacer de ellos las causas verdaderas de los estados de cosas observables que teóricamente, esto es, hipotéticamente, son sus manifestaciones. La psicología cognitiva, en este caso, no se encuentra en desventaja respecto de otras ciencias. Los procesos y estados mentales no observables son las causas verdaderas de las conductas observables que, teóricamente, son sus

manifestaciones. Lo único verdaderamente extraño que parece haber en relación a este punto, es que la psicología cognitiva, al hacer un uso meramente heurístico de la conducta observable, *haya convertido los términos mentales en términos teóricos de la ciencia* y haya transformado, correlativamente, las entidades mentales en entidades teoréticas. Por lo demás, el único problema añadido que se le ha planteado históricamente a la psicología, muy peliagudo por cierto, es el de saber si la única descripción válida de los estados de cosas observables que se toman como índices de los estados de cosas no observables, en tanto que estos últimos son causas postuladas de aquéllos, ha de ser exclusivamente una descripción que finalmente sólo utilice el vocabulario fisicalista. Esta última es la alternativa que adoptó el conductismo metodológico de Hull, Guthrie y Tolman⁴ en la interpretación de la práctica psicológica y que promulgó normativamente para todas las ciencias especiales el fisicalismo del Círculo de Viena. El lema aplicable en este caso diría: si no hay descripción fisicalista de la conducta, no hay posibilidad de aplicar el método postulacional a la psicología y nadie que se dedique en serio a la psicología puede vivir a gusto si esto último no es posible. Pero si es verdad, como he sugerido más arriba, que es característico de la psicología cognitiva la enorme amplitud y diversidad de conductas y fenómenos psicológicos que trata de explicar con el método hipotético-deductivo acuñado sobre modelos lingüísticos y representacionales, entonces la expansión creciente de las conductas sobre las que sería deseable tener una explicación psicológica acertada hace que los férreos corsés explicativos metateóricos que los psicólogos profesionales adoptaron durante el nacimiento y desarrollo del

⁴ Un análisis más detallado de las tesis del conductismo metodológico se puede encontrar en Juan B. Fuentes, 1991. "Un caso ejemplar de historia interna en psicología: continuidad entre la crítica conductista radical del conductismo metodológico y la crítica gestaltista del estructuralismo." El Basilisco (2ª época) 8: 19-39.

conductismo no acierten a contener dentro de sus estrechos límites las explicaciones que provisionalmente parecen imponerse para dar cuenta de aquellos fenómenos. Es de esta manera como se hace enteramente forzoso repasar las tesis fundamentales del fisicalismo si es que se quiere entender algunos aspectos esenciales de la metodología que ha solidificado en la psicología cognitiva. También es esencial realizar esta tarea si se quiere entender qué ha sucedido en filosofía de la mente después del conductismo.

& 4. El fisicalismo en psicología.

Creo que la gran diferencia metodológica entre la psicología cognitiva y la psicología del conductismo metodológico estriba esencialmente en el intento, por parte de aquélla, de abolir el prejuicio fisicalista del ámbito de la metodología psicológica. En este sentido, los psicólogos cognitivos no son, en cuanto a sus aspiraciones metodológicas, la "última generación" de psicólogos conductistas, siquiera sea por el posible cambio conceptual de algunos términos teóricos esenciales. Esta diferencia se puede formular de la siguiente manera: aunque la conducta desplegada por los organismos sujetos a estudio y observable intersubjetivamente es la base empírica de la psicología, no se supone que tal conducta se pueda ni deba explicar y describir finalmente de una manera fisicalista, porque así se dejarían al margen un gran número de fenómenos psicológicos acerca de los cuales nos gustaría tener un modelo explicativo adecuado.

El fisicalismo, como doctrina metacientífica oficial del Círculo de Viena, fue para el conductismo metodológico un lastre pesado del que no consiguió desembarazarse definitivamente y que nubló casi completamente el sentido metodológico y la interpretación de sus investigaciones. Si bien es cierto que el fisicalismo no tuvo tanta repercusión en el dominio de las otras ciencias especiales, sí que influyó decisivamente, metodológicamente

hablando, en el desarrollo de la ciencia psicológica. La física, por ejemplo, contaba con una larga historia de investigación teórica y experimental bien afianzada. El fisicalismo fue sólo un añadido insustancial por lo que hace al curso de esta sólida investigación ya en marcha. Además, la física salió rotundamente fortalecida por el fisicalismo. La unidad de la ciencia, ideal del programa fisicalista, era la unidad, traducción y reducción de todas las ciencias especiales a la física básica.

El fisicalismo, en resumen, fue la tesis de una determinada fundamentación de la ciencia natural y de la descripción del método que debía guiar la investigación de las ciencias naturales, pero no la ciencia natural misma. Exactamente formulado, el fisicalismo no fue, ni es, una teoría acerca de un dominio particular de objetos y, en el caso de la psicología, fue, ante todo, "una teoría lógica acerca de los enunciados de la psicología científica" (Hempel 1972, 128).

En su relación con el fisicalismo, la psicología se encontraba reconocidamente en desventaja respecto de la física, porque creía no ser una ciencia natural y aspiraba denodadamente a serlo. El fisicalismo, en tanto que doctrina metacientífica oficial en los años en que la psicología se hizo consciente de esta aspiración, tuvo un doble efecto en ella. El primero, positivo, fue el de hacer posible que los psicólogos llegaran a percatarse de cuál era la situación exacta de la psicología en el reino de las ciencias naturales. El pronóstico fue desolador: la psicología no es parte de este reino y debe consolarse con pertenecer al de las ciencias del espíritu. Digo "consolarse" porque los juicios de valor empezaron a teñir—no sin razones teóricas de peso—la distinción entre ciencias naturales y ciencias del espíritu. El segundo efecto, negativo, fue que el modelo de investigación al que volvió su mirada la naciente psicología conductista en su anhelo de cientificidad, era demasiado restringido e irreal. Este método fue el hipotético-deductivo de orientación fisicalista. De esta manera, lejos de ser

el fisicalismo un añadido irrelevante para la investigación psicológica, fue absolutamente esencial por lo que hace al sesgo que adquirió la interpretación que los psicólogos dieron al objeto de sus investigaciones y a los resultados que arrojaba el método que estaban utilizando. El fisicalismo se convirtió para la psicología en una teoría normativa. No queda más remedio, por consiguiente, que echar una ojeada retrospectiva al fisicalismo como doctrina de la ciencia y, en particular, como doctrina de la ciencia psicológica, para comprender mínimamente el estado actual en la filosofía de la psicología. Para ello me centraré en las formulaciones del fisicalismo más originarias, las del Círculo de Viena, y no en las más recientes, herederas y continuadoras legítimas de aquél y que se llaman genéricamente, sutilezas aparte, "teoría de la identidad mente-cerebro". El fisicalismo de los años cincuenta de U.T. Place, Herbert Feigl y J.J.C. Smart se puede considerar, sin serias dudas, como heredero legítimo del positivismo fisicalista del Círculo de Viena.

En su aplicación a la psicología, la tesis del fisicalismo, como teoría metacientífica general, dice así: los enunciados psicológicos, tanto los de la vida ordinaria como los de la psicología científica, se refieren a los estados y sucesos físicos que tienen lugar en el cuerpo del organismo cuya conducta se está investigando (Carnap 1965, 203). De forma más condensada y directa, hacer esta afirmación es lo mismo que sostener que "la psicología es una rama de la física" (Ibid.).

Desplegar en su alcance concreto esta tesis obliga a dividirla en dos partes de igual importancia. La primera, que podría llamarse "fisicalismo de abajo arriba", consiste en la afirmación de que la introducción y formación de todos los enunciados y términos de la teoría psicológica debe realizarse sobre la base de enunciados y términos que describen y se refieren finalmente a estados de cosas físicos públicamente observables. En su formulación más extrema, lo que se está sosteniendo aquí es que no es

legítimo que haya una introducción primitiva de enunciados que se refieran a estados de cosas no observables, si es que estos enunciados no se pueden reducir y traducir, por su parte, a enunciados observacionales fisicalistas. La segunda parte, que podría denominarse, correlativamente, "fisicalismo de arriba abajo", es la tesis inversa, según la cual todos los enunciados y términos de la teoría deben ser definibles nuevamente bien lógicamente, bien operacionalmente, en términos de enunciados observacionales que se refieren a estados de cosas físicos del organismo.

Cuando los conductistas metodológicos aplicaron este esquema explicativo al caso de la psicología resultó que el giro fisicalista del método postulacional obligó a describir la conducta--base empírica de la psicología--de forma tal que no se hiciera uso de otro vocabulario que el fisicalista, y a creer firmemente que ese vocabulario era suficiente para describir la conducta que debía ser explicada. Es esta última creencia la que condujo a hacer afirmaciones como ésta: "todas las proposiciones de la psicología describen acontecimientos físicos, a saber, la *conducta física* de los humanos y de otros animales" (Carnap 1965, 171; el subrayado es mío). La intención última del fisicalismo es, pues, despojar a las expresiones psicológicas del uso referencial propio que creemos ingenuamente que tienen; ellas no se refieren a estados psicológicos, sino a estados físicos. Quedaba así cerrada--o se hizo muy complicada--la posibilidad de investigar la conducta intencional. ¿Cómo se describe fisicalistamente la conducta que consiste en la resolución del problema de la Torre de Hanoi? Parece que sólo cuando se expande qué ha de entenderse por conducta--más allá de la idea de "conducta física"--, es cuando se puede tener alguna esperanza de contar con una teoría psicológica acerca de la conducta inteligente. Pero cuando se extiende aquello que ha de incluirse bajo el rótulo de "conducta", el psicólogo tiene que ampliar inevitablemente también el vocabulario observacional, si es que se quiere captar algunas generalizaciones psicológicamente interesantes que

sean válidas para las conductas así descritas. El objetivo del siguiente párrafo será, precisamente, el de recopilar algunos buenos argumentos en contra de la tesis de que la conducta es conducta física, movimiento muscular, estado físico del organismo. Antes, sin embargo, conviene desarrollar un poco más por extenso las dos vías del fisicalismo, para que se vea la fuerza de los argumentos que se le oponen, así como su propia fuerza innegable.

Recorrer el fisicalismo en su vía descendente quiere decir lo mismo que adherirse al verificacionismo como teoría del significado. Para el caso de la psicología, el verificacionismo afirma que una proposición psicológica, sea de la psicología popular, sea de la psicología científica, tiene sentido cuando es posible deducir a partir de ella proposiciones del lenguaje observacional fisicalista. El contenido de la proposición psicológica consiste, precisamente, en que sea posible semejante deducción. Por consiguiente, siempre será posible hacer una traducción o reformulación de cualquier proposición psicológica que tenga sentido a una o varias proposiciones del lenguaje fisicalista que se refieran a los estados físicos del organismo. Se asegura que el contenido significativo de cualquier enunciado psicológico se puede expresar perfectamente bien en términos de enunciados fisicalistas. La seguridad del fisicalismo en sí mismo, su autosuficiencia teórica, se expresa de una forma igualmente contundente en expresiones como éstas:

No debe entenderse que el fisicalismo exija de la psicología que se interese sólo por situaciones físicamente descriptibles. La tesis es, más bien, que la psicología trate de lo que le plazca y formule sus proposiciones como quiera. En cada caso, esas proposiciones serán traducibles al lenguaje fisicalista. (Carnap 1965, 172).

Que los psicólogos profesionales quieran utilizar el lenguaje intencional de las actitudes proposicionales para explicar causalmente la conducta inteligente e intencional está muy bien si es que les es de alguna

utilidad práctica hacerlo así. No importa. Siempre volverán al redil del fisicalismo. Basta con que se persuadan de que ésta es la única vereda racional transitable que es compatible con la metodología científica y, sobre todo, con la ontología materialista, irrenunciable para cualquier psicólogo. Y, efectivamente, los psicólogos conductistas metodológicos se persuadieron completamente de que ésa era la verdadera situación. Las razones eran de peso. Veamos.

Si el objeto de la psicología es la etiología de la conducta de los organismos que el psicólogo observa, entonces esta conducta, antes que nada, habrá de quedar plasmada en enunciados que la describan y se refieran a ella. Según el fisicalismo, estos enunciados son, en primera instancia, enunciados singulares acerca de los estados mentales de los otros; enunciados del tipo "Juan está cansado" o "Juan percibe en el dibujo a una joven con abrigo de piel". En un principio, los fisicalistas fijaron su atención sobre todo en los enunciados singulares que se refieren a los estados psicológicos del otro, y no en los enunciados contrafactuales con la forma de una ley causal, como lo eran aquellos a los que se aludió en el capítulo anterior. Carnap, sin embargo, reconoció estos últimos como la segunda clase fundamental de proposiciones psicológicas generales. La otra clase la constituirían los enunciados acerca de la índole fenomenológica esencial de las distintas especies de vivencias intencionales (Carnap 1965, 174). Un enunciado como (1) "Juan está cansado" forma parte del lenguaje de la psicología de sentido común acerca de la mente y la conducta del otro. "Estar cansado" es un predicado típicamente psicológico. De acuerdo con los principios generales del fisicalismo de "arriba abajo", si este enunciado tiene sentido es porque se puede deducir de él enunciados observacionales en los que sólo aparece vocabulario fisicalista. Sin embargo, curiosamente, Carnap sigue el sentido opuesto del fisicalismo y afirma que ese enunciado tiene sentido cuando se puede inferir a partir de enunciados

observacionales fisicalistas. (Obsérvese lo sospechoso de esta inversión y el aspecto circular a que puede dar lugar adoptar esta construcción metodológica.) En esta inversión, tan característica, consiste el "truco" del fisicalismo. Estos enunciados observacionales fisicalistas, a partir de los cuales se deduce supuestamente el enunciado (1), y que constituyen el contenido semántico propio de éste, serían enunciados como, por ejemplo: (a) "Juan se sienta de una manera peculiar cuando llega a casa", (b) "Juan agacha la cabeza", (c) "Juan no dice nada durante un largo período de tiempo", (d) "Juan resopla de vez en cuando", (e) "Juan sólo mira la televisión", etc. Carnap es consciente, sin embargo, de que semejante conjunción abierta de enunciados fisicalistas no expresaría completamente el contenido del enunciado "Juan está cansado", a no ser que se les uniera la premisa mayor: "cuando percibo que alguien se sienta así y así, y que agacha la cabeza de tal y cual manera, y que no dice nada durante un largo período de tiempo..., es que está cansado, *ceteris paribus*". Otra forma de expresar esta misma idea es diciendo que el contenido significativo del enunciado "Juan está cansado" no coincide plenamente con el contenido significativo de la conjunción abierta de los enunciados (a), (b), (c)..., "sino que va más allá" de ella (Carnap 1965, 176). Como prueba de este "plus ultra" significativo se presenta el hecho de que no se podría obtener el enunciado (1) por medio de inferencias a partir de los enunciados (a), (b), (c)..., si es que a éstos no se les une la premisa mayor.

Es de absoluta importancia crítica retener que el contenido de esta premisa mayor no puede expresarse, por su parte, de una manera completamente fisicalista. Precisamente, la premisa mayor funciona de puente de unión entre el contenido semántico del enunciado psicológico y el contenido semántico de los enunciados observacionales fisicalistas. El antecedente de la premisa mayor contiene predicados fisicalistas, y su consecuente, el predicado psicológico. Es decir, que la traducción completa

de un enunciado psicológico a uno o varios que contienen sólo términos fisicalistas y, de aquí, la supuesta reducción de la psicología a la física, puede realizarse exclusivamente a través de un enunciado condicional que sirve de puente de unión entre ambos y que mezcla ambas clases de vocabularios; el psicológico y el fisicalista. La cuestión difícil es, sobre todo, cómo habría que interpretar la función que esta premisa mayor tiene dentro del proyecto global de reducción del positivismo lógico para el caso de la psicología. Sea como sea el modo en que haya que contestar a esta cuestión, lo que no parece que pueda decirse en ningún caso es que la respuesta se obtenga por mera deducción exclusivamente a partir de los enunciados observacionales fisicalistas. El motivo es fácil de comprender. Desde el momento mismo en que el contenido del enunciado (1) "va más allá" del contenido resultante de la conjunción abierta de enunciados fisicalistas, el movimiento de pasar por inferencias lógicas desde el contenido de los enunciados fisicalistas al contenido del enunciado psicológico, está siempre prohibido o, cuanto menos, es muy problemático. Es decir, la conducta y el estado psicológico de Juan que describimos como "Juan está cansado" no se puede describir completamente—no se puede describir como conducta—apelando exclusivamente a enunciados fisicalistas, sino que hace falta cierto "plus", otra clase de vocabulario al menos. Expresado en el modo material de hablar, para muchos pensadores este plus necesario quiere decir tanto como que el estado de cosas al que se refiere el enunciado (1) no puede reducirse al estado de cosas al que se refieren conjuntamente los enunciados (a), (b),(c), etcétera: a la conducta física. Con todo, antes de proceder de una manera precipitada sobre estas cuestiones, preguntemos, con Carnap, si es que podría haber algo más en la "conducta física"—algo psicológico—que aquello a lo que se refieren los enunciados fisicalistas anteriores.

Carnap augura sombríamente que no, y lo va a demostrar presentando un nuevo conjunto de enunciados que tienen el mismo contenido que (1), es decir, un conjunto de enunciados que, a diferencia de los enunciados fisicalistas (a), (b), (c), etcétera, no necesitan esta vez de una premisa mayor híbrida, como la que antes se requería para establecer relaciones de reducción y traducción entre los predicados psicológicos y los fisicalistas. Este conjunto de enunciados se refiere ahora a ciertos estados y procesos neurofisiológicos que tienen lugar en el cuerpo del organismo cuya conducta se estudia, o que tienen como causa procesos neurofisiológicos. ¿Qué otra cosa podrían ser, si de lo que se trata es de explicar las causas de la "conducta física" de los organismos? No se conoce ningún cuerpo que carezca de sistema nervioso del que estemos dispuestos a admitir que se comporta y, por tanto, que lo hagamos objeto de estudio de la psicología. Hasta donde sabemos, ni las lámparas ni las llantas de los coches tienen conducta.

En la conjunción de enunciados que tienen el mismo contenido que (1), que no quedan "más acá" del contenido de (1), se deberían encontrar, entre otros, "el cuerpo de Juan tiene un pulso lento", "respira también lentamente", "responde de una forma lenta a ciertos estímulos auditivos", etc. Para Carnap, el contenido del enunciado psicológico "Juan está cansado" y el contenido de estos últimos enunciados fisicalistas sí que es ahora idéntico (1965, 178). Hempel, por su parte, afirma lo mismo cuando dice que los enunciados psicológicos son una simple abreviatura de expresiones que serían de por sí demasiado largas y complicadas—la conjunción de enunciados neurofisiológicos—, y cuya especificación completa está lejos de conocerse debido al estado de desarrollo en que se encuentra la ciencia neurofisiológica. Con todo, la expresión psicológica simple y la expresión neurofisiológica complicada tienen el mismo contenido (Hempel 1972, 121). Y ambas tienen el mismo contenido porque ambas son

verdaderas o falsas en razón de las mismas condiciones de satisfacción. Ocurriría aquí, por poner un ejemplo manido, lo mismo que sucede entre el enunciado "cantidad de fluido calórico" y "cantidad media de energía cinética molecular". Es decir, que el enunciado (1) se puede traducir al conjunto de enunciados fisicalistas presentados, porque ambos son verdaderos bajo las mismas condiciones. La ventaja fundamental de este conjunto de enunciados es que ahora contamos con una forma canónica de expresar (1) en la que no aparece ningún predicado psicológico. Carnap manifiesta abiertamente que esta *identidad de contenido* es la que aceptan como principio metodológico fundamental los conductistas metodológicos—a los que él llama "conductistas radicales"—, pero la que rechazan las demás especies de psicólogos. Estos últimos—entre los que habría que contar, supongo yo, a Sellars y a otros tantos— sólo están dispuestos a admitir que todos estos estados físicos son sólo "manifestaciones en su apariencia exterior", "expresiones", del estado mental al que se refiere verdaderamente el enunciado "Juan está cansado". Lo que niega directamente Carnap es que haya nada subyacente, oculto, "un poder oscuro", que sustente y avive (causalmente) estas manifestaciones físicas. Cualquier poder oculto de este estilo sería incognoscible. Pero si es cognoscible, es porque se puede verificar. Y si se puede verificar es porque se reduce a estados físicos.

Esta opinión incurre en el error de una hipostatización, de tal modo que se produce una extraña duplicación: al lado o detrás de un estado de cosas cuya existencia es empíricamente determinable, se supone otra entidad *paralela* cuya existencia no es determinable. (Adviértase que nos referimos aquí siempre a una proposición relativa a mentes de otros.) (Carnap 1965, 179).

Como se suele decir, más claro, agua. La aporía ante la que nos encontramos enfrentados si no sostenemos junto con Carnap y Hempel que los enunciados psicológicos tienen el mismo contenido que ciertos conjuntos

de enunciados físicos--las mismas condiciones de verificación que ellos--, es que los enunciados psicológicos no podrían ser confirmados empíricamente de ninguna manera. Lo cual significa, según la teoría verificacionista del significado, que el enunciado psicológico carecería de significado por ser imposible su contrastación empírica. Lo cual significaría, a su vez, que la psicología como ciencia sería un sinsentido. ¿Cómo se podría verificar siquiera uno de sus enunciados? ¿Qué habría que verificar exactamente? Quien quiera admitir causas mentales de la conducta tiene que enfrentarse con esta aporía.

& 5. Críticas al fisicalismo.

(a) Planteamiento del problema.

Casi nadie, antes de la reflexión, acepta las tesis del fisicalismo de buena gana. Los realistas intencionales--la mayoría de todos nosotros--no las aceptan. Hay muchos pensadores que después de la reflexión siguen sin aceptarlas. Sócrates no las aceptaba y Fodor tampoco las acepta. Quizás no todo lo que hay detrás del rechazo al fisicalismo es, después de todo, una cuestión emocional.

¿Qué se puede alegar reflexivamente contra el fisicalismo positivista? Según el fisicalismo positivista, la introducción de enunciados teóricos en psicología se debe apoyar en la descripción fisicalista de los movimientos en los que se materializa la conducta física observable de los organismos. Una vez que se ejecutó este movimiento metodológico en sentido ascendente, restaba recorrer el movimiento de dirección opuesta. Los términos teóricos de la psicología conductista se deben poder definir de nuevo en un lenguaje fisicalista cuyos términos y enunciados se refieran a los movimientos físicos corporales públicamente observables.

Evidentemente, no hay una forma sencilla de eludir, creo yo, la crítica de que este método puede devenir con cierta facilidad en un círculo explicativo desastroso. La psicología conductista, contagiada hasta la médula de fisicalismo, partía de la descripción de la conducta en términos fisicalistas y volvía a la descripción de la conducta en los mismos términos. Sellars puso una vez más el dedo en la llaga al denunciar que el movimiento descendente del fisicalismo no es ni enteramente necesario, ni tampoco aconsejable.

El requisito conductista de que todos los conceptos de la teoría deberían ser *introducidos* en términos de un vocabulario básico que pertenezca al comportamiento público, es compatible con la idea de que algunos conceptos conductistas deben ser introducidos como conceptos *teóricos*. (1972a, 207).

Con esta opinión de Sellars se puede dejar una puerta abierta a la posibilidad de que los enunciados teóricos de la psicología no tengan por referentes últimos estados y procesos neurofisiológicos del organismo. La introducción primitiva de enunciados teóricos para explicar los hechos condicionales del organismo podría hacer, después de todo, que dichos enunciados se refiriesen a estados mentales causalmente activos¹. Frente a esta tesis de Sellars se encuentra, por ejemplo, la consideración de Hempel ya mencionada. Este filósofo positivista pensaba que en numerosas ocasiones

¹ Es interesante no pasar por alto el hecho de que, posteriormente, Carnap corrigiese parcialmente las posiciones metateóricas que había defendido respecto de la psicología. Aunque siguió considerando que el fisicalismo era una tesis irrenunciable, aceptó la introducción primitiva de enunciados teóricos en psicología. Esta concesión le llevó a relajar mínimamente alguna de sus afirmaciones más radicales. Ahora "los enunciados psicológicos...dicen algo acerca del estado físico de la persona en cuestión. <Pero> son diferentes de los enunciados correspondientes que se enuncian en términos de microfisiología o microfísica...por usar el marco conceptual de la psicología en vez del de aquellos otros dos campos mencionados" (1963, 203). Es decir, que, de alguna manera, Carnap reconoció que el marco conceptual de la psicología es, al menos, diferente del de la neurofisiología. La pregunta es: ¿es esta diferencia un mero capricho que está justificado por su mayor comodidad, o tiene realmente un fundamento en las cosas?

se cae, por imprudencia teórica, en la falsa ilusión que consiste en pensar que términos o enunciados que son meramente una formulación simbólica sucinta de otros términos y enunciados, se refieren a especies de objetos distintos de aquellos a los que se refiere la formulación más larga (1972, 126). Realmente, no habría géneros psicológicos, sino sólo físicos. Esta manera de proceder y de ceder ante las ilusiones referenciales no puede nada más que acarrear cuestiones y discusiones enojosas y absurdas acerca de la "naturaleza intrínseca" de esos objetos; en definitiva, cuestiones que tienen su origen último en un *error categorial*. Distintas maneras de referirse a los mismos objetos y estados de cosas se toman como maneras de referirse a objetos y estados de cosas diferentes. Con las creencias no importa que suceda esta clase de confusión referencial. Se puede creer que la estrella matutina no es la estrella vespertina. Con la ciencia, en donde las creencias subjetivas deberían dar paso a los juicios objetivos, no debería suceder.

El punto decisivo del enfrentamiento entre las tesis de Sellars y de Hempel—como ejemplificación de la disputa entre la metodología psicológica cognitiva y la metodología psicológica fisicalista—, es saber si la introducción primitiva de enunciados teóricos en psicología para explicar causalmente las correlaciones funcionales empíricas halladas en el plano observacional—los "hechos condicionales"—hace que estos enunciados se refieran a especies de objetos—en concreto, a géneros naturales o "esencias"—que tienen propiedades que no se pueden reducir a las propiedades de los objetos que se nos dan inmediatamente en el plano observacional fisicalista. Si la reducción no fuera enteramente posible, entonces el camino descendente del fisicalismo desde los enunciados postulados por la teoría psicológica hasta los enunciados observacionales fisicalistas, tampoco sería enteramente transitable. La psicología no se reduciría a la física o a la neurología. El mundo de las sutilezas, pero también del rigor conceptual, se abre ante quienes quieran

adentrarse en la discusión de este problema metateórico y ontológico, por lo demás tan decisivo... y desesperante. Bienvenido al laberinto. En lo que sigue, sólo espero poder ofrecer un buen hilo de Ariadna con el que poder orientarse a lo largo de las múltiples facetas y esquinas con que está hecho.

Ha sido Fodor, aunque desde luego no sólo él, uno de los psicólogos y filósofos que con más insistencia y, a mi parecer, también acierto, ha logrado reunir algunos argumentos dignos de una mención especial por lo que tienen de crítica a la concepción fisicalista estricta de la psicología que defiende que no hay géneros psicológicos, sino sólo físicos (o neurofisiológicos). Ahora no ofrece ningún problema afirmar que el fisicalismo es primariamente, tal y como afirman sus partidarios, una doctrina lógica—en un sentido peculiar del adjetivo "lógico" que lo hace casi sinónimo de "semántico"—acerca del carácter epistemológico de los enunciados psicológicos y del modo en que éstos deben ser confirmados empíricamente. Por lo demás, es fácil comprender que esta doctrina lógico-semántica conduce irremediabilmente a formular ciertas tesis ontológicas radicales acerca del ser psíquico: el ser psíquico es ser neurofisiológico. Fodor tiene un argumento contra la tesis lógica del fisicalismo y otro contra su contraparte ontológica. En lo que sigue, sin embargo, me centraré especialmente en el análisis del segundo argumento y en el de sus consecuencias, y haré alusiones constantes a otras posiciones que, en uno u otro sentido, lo matizan, amplían y discuten². Finalmente, sería interesante contar con una guía mínima de la maraña de preocupaciones

² Una exposición detallada de las críticas a la parte "lógica" del fisicalismo, que en tantos puntos se entrecruzan inevitablemente con las críticas a la parte "ontológica", puede encontrarse, por ejemplo, en Jerry A. Fodor. 1975. The Language of Thought. (Cambridge, Mass.: Harvard University Press) 2-9; C. S. Chihara y J. A. Fodor. 1991. "Operationalism and Ordinary Language: A Critique of Wittgenstein." (The Nature of Mind, Ed. David M. Rosenthal. Oxford: Oxford University Press) 137-150.

que ha atraído la atención de tantos pensadores como se han ocupado de los problemas relacionados con la crítica al fisicalismo estricto.

En su crítica a la contraparte ontológica del conductismo lógico, Fodor se representa a su interlocutor como alguien que sostiene que todas las psicologías mentalistas confunden desafortunadamente lo que es una relación lógica de inferencia entre enunciados fisicalistas y enunciados psicológicos, sean científicos o de sentido común (Carnap), o una relación semántica de abreviatura de éstos respecto de aquéllos (Hempel), con una relación causal entre el estado de cosas (mental) al que se refiere el enunciado psicológico y los estados de cosas neurofisiológicos a los que se refieren los enunciados fisicalistas. Si los enunciados psicológicos son una forma lingüísticamente sucinta de formular complejas y larguísimas expresiones fisicalistas, entonces no tiene ningún sentido considerar al hipotético estado de cosas al que se referiría el enunciado psicológico como causa de los estados de cosas fisicalistas. La abreviatura y la expresión compleja se refieren inadvertidamente al mismo estado de cosas, y no a dos estados de cosas distintos entre los que existen relaciones causales.

En la crítica de Fodor, sin embargo, el interés no se centra en la refutación de la posibilidad del análisis lógico del significado de los enunciados psicológicos particulares que versan acerca de los estados mentales y el comportamiento de los otros. Antes bien, la compleja argumentación crítica trata primariamente del análisis de las relaciones de reducción de las leyes psicológicas a las leyes físicas (o neurofisiológicas). Con esta última tesis reduccionista, como se puede imaginar, también está comprometido el fisicalismo positivista. Por consiguiente, este giro aparentemente nuevo que tomará el tema de la discusión de aquí en adelante no supone apartarse disimuladamente del problema que se ha suscitado. La razón es aparente y salta a la vista. Si hubiera leyes teóricas psicológicas, los enunciados que las expresarían se tendrían que formular

utilizando predicados psicológicos, los mismos que se utilizan en la formulación de los enunciados psicológicos particulares. Pero si todos los predicados psicológicos no fueran nada más que formas de hablar cómodas acerca de complejos predicados físicos, entonces todas las leyes de la psicología se reducirían finalmente a leyes físicas. Como ya se observó anteriormente, ésta es la tesis fuerte que defiende Carnap. Es por este motivo por el que resulta legítimo decidir la cuestión acerca de la reducción lógica y ontológica--traducción del sentido--de los predicados psicológicos a predicados neurofisiológicos o simplemente físicos--en cualquier caso, siempre fisicalistas--atendiendo principalmente al problema de si sería posible reducir las leyes psicológicas, en cuya formulación tienen que aparecer necesariamente aquellos predicados psicológicos, a leyes físicas, en las que ya no aparecen en absoluto. Si esta reducción fuera posible, entonces habría que negar que haya géneros naturales psicológicos no reducibles a géneros naturales físicos y afirmar, consecuentemente, que, en el fondo, la psicología es física. Por lo demás, no se trata exactamente de decidir ahora mismo si hay verdaderamente leyes psicológicas ni de discutir cuál sería su índole peculiar, sino sólo de saber si se podrían reducir a leyes físicas en caso de que las hubiera. Lo que se discute, más bien, es si hay razones de principio para afirmar esta posibilidad. Hago esta última advertencia por si acaso hay alguien que se echa las manos a la cabeza porque ni siquiera acepte que haya tales leyes psicológicas. La discusión es, por lo tanto, legítima cuando se plantea de esta manera condicional.

La manera condensada que tiene Fodor de plantear todo este variado entrecruzamiento de problemas y de orientaciones en la tradición metacientífica psicológica es sumamente esclarecedora. Supongamos, contrariamente al fisicalismo positivista estricto, que hay legítimas leyes psicológicas que se expresan utilizando predicados psicológicos y no, simplemente, formas cómodas de hablar que utilizan el vocabulario

psicológico y que son útiles para el psicólogo. Bajo esta suposición, el peso de la prueba necesaria para establecer la reducción cae enteramente del lado de los fisicalistas. Para que se pudiera llevar a cabo el programa reduccionista fisicalista, una de las condiciones fundamentales que tendría que darse sería que hubiera leyes puente (bridge laws) entre las supuestas leyes psicológicas y las leyes neurofisiológicas o físicas. La función primordial de estas leyes puente sería la de relacionar el enunciado teórico que aparece como antecedente en la ley psicológica con el enunciado físico que aparece como antecedente en la ley física a la que se debe reducir aquélla. Lo mismo sucede con el enunciado teórico que aparece como consecuente en la ley psicológica. También éste, por medio de una ley puente, se pondría en relación con un enunciado físico que aparecería en la situación de consecuente en la ley física reductora. Dicho brevemente, las leyes puente serían *leyes psicofísicas*. Si hay leyes psicológicas, éstas se reducirán a leyes físicas por medio de leyes psicofísicas. Así, pues, obtenemos la siguiente cadena de expresiones:

$$(1) \quad P1x \rightarrow P2y.$$

$$(2a) \quad P1x \leftrightarrow F1x.$$

$$(2b) \quad P2y \leftrightarrow F2y.$$

$$(3) \quad F1x \rightarrow F2y.$$

La fórmula (1) sería una ley de la psicología, la fórmula (3) una ley de la física o de la neurofisiología—siempre de la ciencia reductora de que se trate—, y las fórmulas (2a) y (2b) serían las leyes puente. Toda la dificultad a partir de ahora se concentra única y exclusivamente en saber cómo deben ser interpretadas las leyes puente y, en especial, cómo debe entenderse el signo lógico " \leftrightarrow " que aparece en su enunciación.

Antes de adentrarse en esta tarea, hay que observar primeramente que la manera ideal de leer la fórmula (1) sería diciendo algo así como que "todos los sucesos que consisten en que x sea $P1$ causan sucesos que consisten en y siendo $P2$ " (Fodor 1975, 10). De una forma exactamente igual habría que leer la fórmula (3), es decir, como un enunciado que expresa una relación causal legal entre los sucesos que consisten en que x sea $F1$ y los sucesos que consisten en que y sea $F2$. Naturalmente, a esta interpretación normal siempre habría que añadirle la famosa cláusula *ceteris paribus* o, lo que es lo mismo a efectos prácticos, hacer una advertencia a propósito del cuantificador universal. Dado que tanto (1) como (3) son leyes naturales, siempre hay que dejar la puerta abierta a las excepciones.

Para hacerse una idea más exacta del alcance de las formulaciones de las leyes puente, es preciso reparar en que la estructura lógica de los enunciados que aparecen a ambos lados de dichas leyes puente es idéntica. En ambos casos se trata de un x o de un y del que se predica un atributo. En concreto, los términos singulares que funcionan como sujetos gramaticales en las leyes puente y sobre los que recaen las predicaciones psicológica y física, son los mismos; a saber " x " e " y ". En general, se supone que el referente último de los términos singulares habrá de ser el organismo corporal completo. Todavía no se ha hecho psicología experimental con ningún sujeto-no-corporal (de haberlo realmente), ni con ningún cerebro aislado en una cubeta. (Para el fisicalista estricto quizás no hubiera ninguna imposibilidad de principio para hacer esto último.) Como ha recordado Nagel "es dudoso que cualquier cosa sin un cuerpo de la clase convencional pudiera ser el sujeto de un estado psicológico" (1970, 217). Creo que es por esta razón, aunque Fodor no lo diga explícitamente, por la que en las leyes puente deben aparecer los mismos sujetos gramaticales como sujetos, unas veces, de predicados psicológicos y, otras veces, como sujetos de predicados físicos.

(b) Complicaciones ontológicas de toda posible superación del fisicalismo positivista en psicología.

Hechas estas advertencias de rigor, Fodor procede a enfrentarse con la difícil interpretación del signo de las leyes puente. Como es fácil prever, para el programa fisicalista estricto no habría duda alguna de que este signo se debe interpretar sin excepciones como un signo de *identidad* estricta entre sucesos. Los sucesos que satisfacen el antecedente de las leyes puente son idénticos con los sucesos que satisfacen su consecuente. Por lo tanto, desde el momento mismo en que los sucesos a los que se aplican los predicados psicológicos son sucesos físicos, no hay en los organismos sucesos psicológicos que pertenezcan a un género diferente de los sucesos que pertenecen a los géneros físicos. Por otra parte, dado que, como consecuencia de esta tesis, todos los géneros son géneros físicos, en el fondo no habría leyes psicofísicas, porque toda ley psicológica es, en realidad, una ley física.

¿Cabe, sin embargo, otra interpretación de las leyes puente que sea distinta de su interpretación como enunciados de identidad? Sí, dice Fodor. La otra posibilidad de interpretación de las leyes puente, la posibilidad contraria al fisicalismo estricto--y a partir de aquí comienzan las imprescindibles y escurridizas sutilezas anunciadas--, es considerar que el signo de las leyes puente expresa *relaciones causales*. Debería tener, pues, la misma interpretación que el signo que aparece tanto en la expresión de la ley reductora (3) como en la que hay que reducir (1). En esta ocasión, lo que diría una ley puente es que, por ley, el suceso que consiste en que *x* satisfaga un predicado psicológico está causalmente correlacionado con el suceso que consiste en que *x* también satisfaga un predicado físico. Por otra parte, dado que el objeto sobre el que recaen las predicaciones en las leyes puente es el mismo *x*, lo que dice una ley puente es que es

nomológicamente necesario que los predicados *P* y *F* se apliquen a las mismas cosas (Fodor 1975, 12). Pero decir esto último, según Fodor, "es compatible con una ontología no fisicalista, dado que es compatible con la posibilidad de que el hecho de que *x* satisfaga *P* no sea por sí mismo un suceso físico" (Ibid.). Merece la pena seguir leyendo las palabras de Fodor que conducen a su tesis ontológica fundamental más original (¿cercana, después de todo, a la de Spinoza?):

Se podría imaginar, por ejemplo, una doctrina de acuerdo con la cual los predicados físicos y psicológicos se apliquen a los organismos, pero en la que se niegue que el suceso que consiste en que un organismo satisfaga un predicado psicológico sea, en cualquier sentido, un suceso físico. El resultado sería una especie de dualismo psicofísico de una variedad no cartesiana: un dualismo de sucesos y/o propiedades, en lugar de un dualismo de sustancias. (1975, 12).

De ser verdadera, la consecuencia más importante e inmediata de esta tesis sería que la psicología como ciencia no podría reducirse nunca a la física básica ni a ninguna otra ciencia especial, porque hay sucesos que satisfacen predicados psicológicos, pero que no satisfacen predicados físicos, aun cuando un suceso y otro se den en, o lo sean de, los mismos objetos: los organismos. (Se entiende en todo momento que un suceso es un hecho individual sometido al tiempo.)

En este caso y por lo que respecta exclusivamente a esta circunstancia, estamos ante una situación muy semejante a la que se da cuando decimos que el suceso que consiste en que *x* sea de este rojo es distinto del suceso que consiste en que *x* tenga esta extensión, aunque uno y otro suceso lo sean de, o se den en, *este* objeto singular. Esto último se debe a que en los objetos físicos singulares no puede haber casos particulares de color que sean, en la terminología de la ontología formal fenomenológica, partes independientes o trozos de casos particulares de

extensión. Sin embargo, como se podrá apreciar en breve, en el caso de la relación entre los sucesos psicológicos y los sucesos físicos la situación no es, ni con mucho, tan sencilla. De momento, a casi nadie se le ha ocurrido reducir la ciencia de los colores--aquella que versa sobre los sucesos que satisfacen predicados de color--a la ciencia de la extensión--aquella que versa sobre los sucesos que satisfacen predicados de extensión--, cosa que sí ha sucedido y sigue sucediendo en el caso de la ciencia psicológica y de la ciencia física o neurofisiológica. Otra razón de peso se puede adelantar ya. Mientras que, por una "ley de esencias", como se suele decir, no puede suceder que haya casos particulares de color sin que haya también casos particulares de extensión--ambos sucesos son necesariamente coextensivos espacio-temporalmente--, sí que puede haber casos particulares de sucesos físicos sin que haya casos particulares de sucesos psicológicos. Las piedras satisfacen, con toda seguridad, predicados físicos, pero no psicológicos. Y, así, la ley que correlaciona espacio-temporalmente sucesos físicos y psicológicos tiene que ser una ley diferente de una "ley de esencias". Quizás sea una ley empírica.

Por lo pronto, el cuadro ontológico sintético que propone Fodor es pensar (i) que en un mismo objeto material--sin decir aquí nada todavía acerca de qué es esencialmente el ser material--se dan casos particulares de diversas especies de sucesos; esos sucesos satisfacen unas veces predicados físicos y otras predicados psicológicos, y (ii) que entre unos sucesos y otros se dan relaciones causales. La ventaja de esta posición ontológica respecto de la del fisicalismo estricto es que con ella no parece necesario salir todavía del círculo de la ontología materialista, aunque se reconozca, por otra parte, la peculiaridad de las propiedades psicológicas de los organismos. Gracias a este dualismo de propiedades, que no de sustancias, se puede seguir siendo materialista sin ser ya fisicalista acérrimo. (Para una postura semejante véase Searle (1984a, 6).)

Fodor ha definido en otro lugar qué entiende exactamente por dualismo de sustancias y en qué sentido no es preciso inscribirse necesariamente en esta posición ontológica, aunque se defienda el dualismo de las propiedades psicológicas y físicas. Sólo una aclaración terminológica previa. En esta ocasión ha llamado simplemente "dualismo" al dualismo de las sustancias, y ha reservado exclusivamente el rótulo de "mentalismo" para la ontología que defiende el dualismo de las propiedades psicológicas y físicas, pero el monismo materialista de las sustancias.

Estoy usando el término "mentalismo" de forma que se hace necesario distinguir entre mentalismo y dualismo, distinción que no siempre se hace en la literatura filosófica. Entiendo por "dualismo" la doctrina de que los predicados mentales y conductuales <físicos> se aplican a sujetos de tipos lógicos esencialmente diferentes—es decir, que los predicados mentales designan estados de la mente, siendo ésta una sustancia diferente, de modo irreductible, del cuerpo, al cual se aplican los predicados conductuales. En general, se podrá llamar dualista a aquel que sostenga que tanto la mente como el cuerpo son cosas y que son cosas diferentes

.... Aunque es normal que todos los dualistas sean mentalistas, lo contrario no es, de manera alguna, verdadero. Afirmar que los juicios acerca de la mente y los juicios acerca de la conducta son lógicamente independientes, es algo plenamente compatible con afirmar que los predicados mentales y conductuales se aplican al mismo tipo de sustancia—como a las personas y a los objetos físicos, por ejemplo. Llegamos aquí a un tercer aspecto en el que la terminología que usamos se aparta de lo que es tradicional en la literatura filosófica...: que el mentalismo, tal y como lo hemos definido, es compatible con la definición tradicional de materialismo. En concreto, es compatible con la doctrina de que los estados de la mente y los estados del cuerpo son *idénticos de una manera contingente*. (1980b, 88-89; el subrayado es mío).

Si es importante llamar la atención sobre este texto es porque en él se introducen dos piezas ontológicas nuevas que son fundamentales para la discusión. Primeramente, se presenta la idea de que los estados psicológicos y los estados del cuerpo son *contingentemente idénticos*, y, en

segundo lugar, la idea de que el mentalismo—la defensa conjugada del monismo materialista de las sustancias y del dualismo de las propiedades psicológicas y físicas—está comprometido con esta identidad contingente de los sucesos psicológicos y físicos. Pero, ¿qué quiere decir propiamente esta idea extrañísima de la identidad contingente entre sucesos? ¿Acaso se puede distinguir ésta de la identidad estricta que postula el fisicalismo positivista? ¿Se pueden hacer, entonces, distinciones específicas dentro del género identidad? Y, sobre todo, ¿a partir de qué clase de reflexiones se podría llamar a la idea de la identidad contingente entre sucesos para hacer de ella un paso necesario en la solución del problema de la interpretación causal de las leyes puente? Lo que quiero decir con esta última pregunta se hace patente si se considera que en ningún caso nos vemos obligados a afirmar, por ejemplo, que el ser rojo de este objeto sea un suceso contingentemente idéntico con el ser extenso de este mismo objeto. También podemos defender el dualismo de las propiedades y el monismo de las sustancias para esta última clase de situación ontológica, sin necesidad de hablar en ningún caso de sucesos contingentemente idénticos, sino sólo de sucesos que no son troceables y que son coextensivos espacio-temporalmente. Se necesita, pues, algún as que no está claramente todavía encima de la mesa, aunque no esté tampoco enteramente ausente. ¿Cuál podría ser esta carta escondida?

Antes de intentar dar respuesta precisa a este interrogante, conviene resumir lo que tenemos hasta el momento. Si no queremos interpretar las leyes puente como enunciados que expresan la identidad estricta entre la clase de los sucesos que satisfacen el antecedente de las leyes puente con la clase de los sucesos que satisfacen su consecuente—negando, por consiguiente, la causación psico-física, la existencia de leyes psicofísicas y la peculiaridad ontológica del ser psíquico—, entonces, según Fodor, debemos interpretarlas, en un principio, como enunciados que expresan correlaciones causales legales. Según esta segunda interpretación, estamos comprometidos

a sostener diversas tesis. Las principales son: (1) los sucesos que consisten en que x satisfaga un predicado psicológico P son diferentes de los sucesos que consisten en que x satisfaga un predicado físico F ; (2) que dado que es el mismo x , el organismo sujeto de experimentación psicológica, el que satisface tanto predicados P como predicados F , los sucesos que consisten en que x satisfaga P habrán de ser contingentemente idénticos con los sucesos que consisten en que x satisfaga F ; (3) que entre los sucesos psicológicos del organismo y los sucesos físicos se dan correlaciones causales.

Como se puede apreciar sin gran esfuerzo, defender la segunda versión de las leyes puente complica las cosas extraordinariamente, aunque nadie ha dicho nunca que las cuestiones de filosofía de la psicología tengan que ser necesariamente sencillas. Las opciones son, por tanto, limitadas: o complicarse la vida, o bien decir que todos los sucesos del universo son simple y llanamente sucesos físicos. Si uno elige complicarse la vida--y lo harán todos aquellos que no creen defendible hasta el final el fisicalismo estricto--, entonces no podrá hacer otra cosa que complicar también su modo de expresión. Una muestra: Fodor opta por complicarse la vida, y por este motivo se puede leer en sus escritos cosas tan curiosas como ésta: las leyes puente enuncian "identidades contingentes entre sucesos que son nomológicamente necesarias" (nomologically necessary contingent event identities) (Fodor 1975, 13). Además de ésta, se pueden leer otras expresiones que tienen todo el aire de ser igualmente contradictorias. Algunas acaban de ser reflejadas en varios momentos: "...pero en la que se niegue que el suceso que consiste en que un organismo satisfaga un predicado psicológico sea, en cualquier sentido, un suceso físico" (1975, 12); "las leyes puente como la fórmula (2) podrían ser aceptadas como la expresión de identidades contingentes entre sucesos" (1975, 12); "los enunciados puente expresan identidades-muestra (token identities) entre

sucesos" (1975, 22). Esta serie de expresiones no son otra cosa que otras tantas manifestaciones de la defensa de que los sucesos que consisten en que x satisfaga un predicado psicológico son diferentes de los que consisten en que x satisfaga uno físico, aunque, por otra parte, ambos sucesos sean contingentemente idénticos. Lo que está pidiendo a gritos la teoría de Fodor, desde mi punto de vista, es una explicación algo más detallada de cómo es posible afirmar estas dos cosas a la vez.

Antes de abordar esta dificultad, quiero hacer una indicación "histórica". Fodor no está solo en medio de toda esta polvareda de complicaciones. Un camino semejante en muchos aspectos al que él ha transitado es el que ha defendido también, por ejemplo, Donald Davidson, aunque este último haya subrayado aún mucho más el carácter "aparentemente contradictorio" de las tesis expuestas y que él resume como la defensa de estos tres principios tan conocidos (1979, 80-81):

(i) El principio de la interacción causal entre sucesos psicológicos y físicos. Sostenemos que los sucesos mentales interactúan causalmente con los sucesos físicos, y viceversa.

(ii) El principio del carácter nomológico de la causalidad: donde hay causalidad entre sucesos debe haber también leyes estrictamente deterministas.

(iii) El principio del anomalismo de lo mental: no hay leyes estrictamente deterministas sobre cuya base se puedan explicar y predecir los sucesos mentales.

La tesis de Davidson es que se puede poner en perfecta armonía estos tres principios si es que se defiende cierta versión de la teoría de la identidad entre lo mental y lo físico (1979, 81). La pregunta es ahora, como lo era en el caso de Fodor, saber cuál es esta versión que disipa la aparente contradicción que se encierra en los principios mencionados. A partir de este punto, los razonamientos de Fodor y Davidson corren básicamente al unísono. El argumento surge a partir de ciertas reflexiones

que son relativamente independientes de aquellas que se han estado considerando hasta este momento y que tienen que ver con el modo en que se realiza normalmente la taxonomía de los géneros naturales. Por consiguiente, es importante estar atentos a este argumento porque él nos dará el elemento necesario que estaba haciéndose esperar desde hace tanto tiempo para apuntalar la inteligibilidad del concepto fodoriano de la identidad contingente entre sucesos o, en el caso de Davidson, entender cómo es posible que haya causación psicofísica, aunque no haya leyes psicofísicas.

(c) Inconmensurabilidad nomológica entre sucesos.

Para ver con más detalle cómo sería posible diluir la sensación de contradicción que pueden transmitir las teorías de Fodor y Davidson, y mostrar que estamos ante una situación ontológicamente factible, si bien sumamente curiosa, es preciso destacar algunas consecuencias que estaban contenidas implícitamente en algunos puntos que se trataron en el capítulo precedente. Me refiero de nuevo a la necesidad de contar con una descripción provisional de las propiedades de los objetos que caen dentro del dominio de las ciencias especiales y que permitiría la primera taxonomía provisional de los mismos. Como se argumentó entonces, esta primera descripción no asegura de antemano que los sucesos así descritos lleguen definitivamente a formar parte de la ciencia madura idealmente completada. O, lo que quiere decir lo mismo, que no hay garantías de que dichos sucesos caigan bajo las leyes propias de dicha ciencia cuando se describen de esta manera previsiblemente transitoria. La falta de garantías en este punto es lo que permitió la distinción entre el fundamento de la inteligibilidad de una ciencia y su fundamento ontológico. Estar abiertos al ser no es conocer el ser. Sin embargo, lo que es posible prever fiablemente

es que las descripciones que la ciencia idealmente completada proporcionaría de sus objetos sí que serían descripciones que satisfacen predicados de géneros naturales; sus predicados se referirían a géneros naturales y sus leyes subsumirían sucesos que, bajo esas descripciones, satisfarían esos predicados. Este es, precisamente, el fin último de la ciencia, su idea misma: determinar los géneros naturales, los tipos de entidades cuyas descripciones satisfacen clasificaciones científicamente relevantes.

La cuestión se centra ahora en averiguar si contamos de momento con alguna razón de principio para sostener que todos los géneros naturales de las ciencias especiales idealmente completadas--entre las que se encontraría la psicología--serán finalmente géneros físicos. Jerry Fodor (1975, 15) y, antes que él Thomas Nagel (1970, 222-223), pero también Donald Davidson (1985, 110-111) y Jaegwon Kim (1979, 40), por ejemplo, consideran que no.

La razón fundamental para sostener esta tesis crucial es que puede haber interesantes generalizaciones causales--generalizaciones sustentadas por enunciados contrafactuales--acerca de sucesos cuyas descripciones y propiedades físicas no tienen rasgos comunes uniformes que sean relevantes para su taxonomía científica, aun cuando todos estos sucesos caigan precisamente bajo aquellas generalizaciones científicamente interesantes. Por lo demás, aunque los sucesos que caen bajo estas generalizaciones especiales tuvieran, de hecho, descripciones físicas en común, esta circunstancia sería "enteramente irrelevante para la verdad de tales generalizaciones" (Fodor 1975, 15). Como se argumentará correlativamente para el caso de la psicología, que haya una correlación causal entre sucesos que son la instanciación de una ley psicológica general, es compatible con el hecho de que los sucesos así correlacionados tengan distintas descripciones físicas, descripciones físicas disyuntas.

El ejemplo analógico que propone Fodor para ilustrar esta posibilidad descansa en la importancia que tiene tomarse absolutamente en serio la

existencia de ciencias especiales y de instituciones científicas relativamente independientes. Así, por ejemplo, si la ley económica de Gresham acerca de qué sucede con los intercambios monetarios en ciertas condiciones fuese verdadera, entonces cualquier suceso que cayera bajo esta ley sería compatible, en principio, con una disyunción indefinida de descripciones físicas de los sucesos que constituyen, de hecho, intercambios económicos. Fodor está dispuesto a admitir, con todo, que "cualquier suceso que consista en un intercambio monetario (y, de aquí, cualquier cosa que caiga bajo la ley de Gresham) tiene una descripción verdadera en el vocabulario de la física en virtud de la cual cae bajo las leyes de la física" (1975, 15; véase también Davidson (1985, 110-111) y Pylyshyn (1987, 108)). Lo que es esencial para el argumento es que no hay *una* descripción física enteramente uniforme que se aplique por igual a todos los sucesos que estamos dispuestos a considerar como intercambios monetarios. La descripción física de ellos ha de ser *disyunta* y abierta. ¿Qué tienen en común, en cuanto a su descripción física, un intercambio de conchas, firmar un cheque o hacer transacciones de billetes de banco? Sin embargo, todos estos sucesos son intercambios económicos y caen, en virtud de esta descripción común, bajo la ley de Gresham, de ser verdadera. En casos como éste se dice técnicamente que los sucesos que constituyen intercambios económicos son "nomológicamente inconmensurables" respecto de los sucesos físicos (Kim 1979, 39). Es decir, que aunque todo suceso económico tenga una descripción verdadera en el vocabulario de la física, es altamente improbable pensar que los predicados de la economía y los de la física tengan la misma extensión y esto último, además, por ley.

Por consiguiente, si es convincente esta serie de ideas, entonces una ley puente que quisiera reducir los predicados de la economía a los predicados de la física tendría una formulación semejante a ésta:

X es un intercambio monetario $\leftrightarrow F1x \vee F2x \vee F3x \vee \dots \vee F'x$

El problema es que, de momento, no hay garantías ni hay que albergar tampoco esperanzas especiales—aunque aquí no se trate, como se acaba de señalar, de necesidades apodícticas, sino de cuestiones de hecho, de hechos acerca de cómo está constituido probablemente el mundo—, para pensar que el consecuente de una ley puente como ésta, por ser una disyunción de predicados, en principio, abierta, se refiera a ningún género natural físico. Sencillamente " $F1x \vee F2x \vee F3x \vee \dots \vee F'x$ " no es un predicado que se refiera a género natural físico alguno. Pero si no se refiere a ningún género físico, es difícil que la ley puente sea efectivamente una ley, es decir, que los predicados económicos sean nomológicamente coextensivos con los predicados físicos, dado que "una condición necesaria para que una generalización universal sea legal, es que los predicados que constituyen su antecedente y su consecuente sean predicados de géneros" (Fodor 1975, 20). Por lo demás, como Jaegwon Kim ha hecho notar, la situación de inconmensurabilidad nomológica entre sucesos es completamente normal tanto en las ciencias como fuera de ellas. Así, por ejemplo, son también nomológicamente inconmensurables las propiedades que normalmente atribuimos a los objetos de tamaño medio que están a nuestro alcance respecto de ciertas otras propiedades que también atribuimos a las microestructuras que constituyen esos objetos. Por ejemplo, no esperamos encontrar ninguna estructura microfísica uniforme que subyazca a todos y sólo los objetos que son mesas (Kim 1979, 39).

Correlativamente con estos ejemplos, no hay que albergar tampoco esperanzas especiales en relación al hecho de que los predicados que expresan géneros psicológicos sean nomológicamente coextensivos con predicados que expresan géneros físicos, aunque todo estado y proceso psicológico particular tenga una descripción física verdadera—es decir, sea

verdad respecto de él algún predicado físico--en virtud de la cual ese suceso cae bajo leyes físicas. La motivación para establecer la plausibilidad de esta tesis descansa en la misma clase de fundamentos que en el caso de la economía. Seguramente es verdad decir de diferentes individuos que están pensando en Venus, o que han pensado en Venus en diferentes ocasiones. Uno de ellos, sin embargo, ha pensado en este planeta como aquel punto luminoso que estaba mirando a la caída de la tarde cuando meditaba sobre sus problemas. Otro, pensaría en Venus cuando, al salir de su casa por la mañana temprano, miró a esa luminosidad en el cielo y exclamó "¡qué madrugada tan bonita!". Otro, pensó en Venus cuando descubrió que la estrella matutina es la estrella vespertina. Pero dado, como es canónico decir, que el tipo de suceso psicológico en que consiste "pensar en Venus" y que así caracterizamos, puede tener asociados diferentes contenidos mentales, entonces *"es altamente improbable que algún estado neuronal uniforme se correlacionará, de una manera uno a uno, con la ocurrencia de este suceso psicológico"* (Kim 1979, 38) Que esto sea así, no quita para que un organismo esté en algún estado físico individual cuando está en un determinado estado psicológico individual. Además de esta clase de argumentaciones para apoyar el punto de vista de la inconmensurabilidad nomológica entre predicados físicos y predicados psicológicos, también se suele presentar como prueba la evidencia empírica de que determinadas estructuras neurofisiológicas idénticas de un organismo pueden cumplir distintas funciones psicológicas en diferentes momentos. Estas dos clases de evidencias y de argumentaciones son las que han conducido clásicamente a la defensa del psicofuncionalismo como doctrina ontológica "oficial" de la filosofía de la mente contemporánea: los mismos estados psicológicos se pueden realizar de una manera múltiple en diversas estructuras físicas de tal forma que resulta de todo punto improbable que cada estado psicológico particular esté correlacionado siempre con un único estado neurofisiológico.

Por ese motivo, aunque dos organismos difieran en sus estados físicos no diferirán por ello necesariamente en sus estados psicológicos. Por ejemplo, es muy improbable que la realización física de mi estado psicológico de dolor o de pena o de creencia sea idéntica a la realización física de ese mismo estado genérico en ti. Otra fuente complementaria de evidencias empíricas normalmente presentada para sostener el psicofuncionalismo, es que posiblemente haya sistemas físicos no orgánicos que satisfacen predicados psicológicos—es razonable hacer de ellos sujetos lógicos de atribuciones de predicados psicológicos—pero que no satisfacen, ni evidentemente pueden satisfacer, predicados neurofisiológicos.

Todas estas comparaciones entre los sucesos que satisfacen predicados psicológicos y los sucesos que satisfacen predicados propios de los dominios de otras ciencias especiales, tenían como finalidad resaltar de qué manera se podría hablar de sucesos contingentemente idénticos para defender así la tesis comúnmente conocida como fisicalismo-muestra (token-physicalism) que establece identidades entre *casos particulares* de estados psicológicos y *casos particulares* de estados físicos. Sin embargo, la clase de reduccionismo que propugna el ideal del fisicalismo estricto positivista que se ha analizado en el párrafo anterior, parece exigir que se dé una correspondencia mucho más fuerte, una correspondencia nomológica entre tipos o géneros de estados psicológicos y tipos o géneros de estados físicos³. Las leyes puente

³ Entre otros lugares, se puede encontrar una discusión detenida de por qué la reducción fisicalista estricta requiere identidad entre tipos de predicados físicos y tipos de predicados psicológicos, y no meramente entre casos particulares, en los siguientes artículos: Laurence Mucciolo. 1975. "Neurophysiological Reduction, Psychological Explanation and Neuropsychology." Philosophy of the Social Sciences 5: 451-462; Ned J. Block y Jerry A. Fodor. 1972. "What Psychological States Are Not." Philosophical Review April 1972: 159-181; Jaegwon Kim. 1979. "Causality, Identity, and Supervenience in the Mind-Body Problem." Midwest Studies in Philosophy, Studies in Metaphysics. Eds. Peter A. French et al. Minneapolis: University of Minnesota Press: 31-49; Donald Davidson. 1970. "Mental Events." Experience and Theory. Ed. Lawrence Foster & J. W. Swanson. Mass.: The University of Massachusetts Press: 85.

son realmente leyes. Esta es la tesis del fisicalismo-tipo (type-physicalism). Sin embargo, con una justificación firme de esta última tesis es con lo que no podemos contar a juzgar por las consideraciones anteriores.

La consecuencia más probable que se extrae de defender la clase de evidencia que habla en contra del fisicalismo-tipo, es que aunque haya relaciones uno a uno entre muestras de estados psicológicos y muestras de estados físicos, no hay relaciones legales nomológicas entre clases de estados psicológicos y clases de estados físicos. En este caso, la tesis de Fodor coincide plenamente con la de Davidson en el sentido de sostener que hay causalidad psicofísica, pero que esta correlación causal no es legal (Davidson 1970, 85). Como se recordará, en esta tesis consiste precisamente para Davidson el "anomalismo de lo mental". Es decir, que, propiamente hablando, no habría leyes psicofísicas.

Una de las clases de evidencia fundamentales para abandonar el fisicalismo-tipo (la coextensión nomológica de los predicados psicológicos con los predicados físicos), en favor del fisicalismo-muestra, acaba de ser tratada y dice que no hay descripciones físicas uniformes de muchas especies de sucesos que caen bajo generalizaciones científicamente relevantes. Además de esta razón, y superponiéndose con ella en varios puntos, se suele ofrecer también otras tres. Las dos primeras razones son clásicas, pero Fodor y Davidson consideran que son, hasta cierto punto, irrelevantes y no del todo buenas. Sólo la tercera sería realmente valiosa⁴. Así, pues, primero la tesis que hay que refutar y después los tres argumentos:

Tesis:

Los predicados físicos y los predicados psicológicos (o de acción) son nomológicamente coextensivos.

⁴ Debo la aclaración y sistematización de este punto a Ramón del Castillo.

Argumentos en contra:

(i) No puede haber predicados físicos coextensivos con predicados de acción, porque una misma acción se puede ejecutar de modos completamente heterogéneos. Por ejemplo, se puede saludar de muy diversas maneras (Davidson 1985, 115; Fodor 1980b, 72).

(ii) No puede haber predicados físicos coextensivos con predicados de acción, porque los mismos movimientos físicos pueden contar como la realización de acciones muy diferentes. Mover la cabeza alternativamente de derecha a izquierda puede contar como un ejercicio gimnástico o como un movimiento suficiente para percibir una habitación por nuevos lados (Davidson 1985, 115).

(iii) No hay ningún movimiento que sea suficiente para la realización de una acción por muy íntima que sea la unión--a veces se habla aquí de *uniones internas*--entre la realización del movimiento y la realización de la acción (Fodor 1980b, 74-75).

Un ejemplo analógico sirve para proporcionar cierta luz a lo que declara este último argumento. Advuértase de paso que si este argumento es válido, entonces se contará también con un argumento que previsiblemente podría emplearse en contra del conductismo lógico, porque lo que él declara es que no hay ninguna relación lógica entre la realización de un movimiento y la realización de una acción. He aquí el ejemplo prometido. Que una pelota de baloncesto atraviese el aro de la canasta de arriba abajo es un suceso necesario para que se marquen puntos en un partido de baloncesto. La unión entre este movimiento de la pelota y marcar dos o tres puntos es tan íntima que nada contará como marcar dos o tres puntos si es que la pelota no atraviesa el aro de baloncesto de arriba abajo. En el sentido de Wittgenstein, éste es uno de los *criterios* que establecen las reglas del baloncesto para decidir cuándo se han marcado

puntos. Sin embargo, que la pelota pase por el aro no es una condición suficiente para aumentar el tanteador de uno de los equipos; un suceso no *entraña* el otro, aunque "la relación entre el criterio y aquello de lo que es criterio sea lógica o conceptual" (Chihara y Fodor 1991, 139). Además de satisfacer esta condición, se requieren muchas otras cosas más. Así, por ejemplo, la pelota tiene que estar en juego, es decir, no tiene que haberse cometido antes una falta; el partido tiene que haber comenzado; tiene que estar corriendo el tiempo de juego efectivo, etcétera. Sin embargo, lo que resulta especialmente relevante para el argumento general es que no hay visos de que se puedan detallar las condiciones bajo las cuales el suceso que consiste en que la pelota pase por el aro cuente realmente como marcar puntos "*únicamente* por referencia a otros movimientos" (Fodor 1980b, 75; el subrayado es mío). Se podría, con todo, intentar hacerlo y decir cosas tales como que el suceso que consiste en que la pelota atraviese el aro de arriba abajo contará como marcar tres puntos si es que la pelota ha sido lanzada desde detrás de la línea de los 6'25 metros. Pero esta condición por sí misma tampoco sirve, porque si un jugador de ese partido lanza la pelota desde cincuenta metros de distancia ese movimiento tampoco contaría como marcar tres puntos. Seguramente habría realizado dicho movimiento con los pies apoyados fuera de la cancha y eso, como los aficionados saben, tampoco vale. La apuesta es que se intente especificar únicamente por referencia a otros movimientos las condiciones bajo las cuales el hecho de que la pelota atraviese de arriba abajo el aro de la canasta cuente como marcar puntos en baloncesto. Los pronósticos apuntan a que así nunca se conseguirá hacerlo. No son más y más descripciones de movimientos físicos, sino las reglas del baloncesto las que lograrían, de una manera peculiarmente simple, establecer estas condiciones. Dicho con la mayor brevedad posible: las reglas son condiciones para que un movimiento, o una serie determinada de ellos, cuente como la realización

de un tanto, y las condiciones que establecen las reglas no se especifican únicamente por referencia a otros movimientos. Las reglas son, por otra parte, condiciones que imponen un determinado tipo de vocabulario para la descripción adecuada del mundo. Las reglas del baloncesto son condiciones cuyo conocimiento permite hablar de marcar puntos en baloncesto y no meramente de que la pelota pase por el aro de arriba abajo. Seguramente, la descripción adecuada del mundo perdería en capacidad, riqueza y exactitud si es que no podemos hablar de marcar puntos en baloncesto. Sin embargo, dice también Fodor en otro lugar que "del hecho de que muchos movimientos diferentes puedan asociarse con la misma acción, no se puede inferir que no exista ningún movimiento cuya ocurrencia sea suficiente para realizar la acción" (1980b, 72-73). Si se aplica esta última afirmación al ejemplo del partido de baloncesto que estamos considerando, la intuición que nos invita a compartir Fodor es que, de todas las maneras, ciertos movimientos han sido, de hecho, la realización plena de un cierto partido de baloncesto⁵.

Consiguientemente, y de una manera correlativa, también estamos invitados a compartir la idea de que podemos tratar a los predicados psicológicos y de acciones respecto de los estados físicos y los movimientos particulares del organismo que los realiza en el mundo físico, como a las reglas del baloncesto respecto de los variadísimos movimientos que siempre habrán de constituir enteramente un partido de baloncesto. La psicología, si se me permite esta metáfora, sería la "regla del cuerpo". (¿No resultará

⁵ Fodor pone otro ejemplo enteramente análogo a éste para ilustrar el mismo punto. En un cierto sentido, la acción que consiste en mover un alfil se puede realizar simplemente cogiendo el alfil con la mano y cambiándolo de lugar. Esta acción, obviamente, se puede realizar además de múltiples formas, dado que los movimientos que conducen a su ejecución son ilimitados. Todo depende de lo grande que sea el tablero, del tamaño y forma de las piezas, etc. Pero en otro sentido, mucho más relevante que el anterior, se considerará que se ha realizado efectivamente la acción de mover el alfil si es que se han respetado íntegramente las reglas del ajedrez (Fodor, 1980b, 74).

que todo lo que hemos estado haciendo inadvertidamente durante este tiempo ha sido dar vueltas alrededor de algunas ideas de origen aristotélico tan antiguas como buenas?)

Después de oír estas razones, alguien podría decir que todo esto tiene un aire muy extraño y "anómalo". Por una parte se nos quiere convencer de la verdad de (iii), de que no hay ningún movimiento que sea suficiente para la realización de una acción, y, por otra, se nos insiste para que aceptemos que tiene que haber un conjunto de movimientos que constituyan completamente la realización de esa acción. Parece que nunca se acaba de ir de aparente contradicción en aparente contradicción. Sin embargo, sea lo que sea de esta cuestión, no hay que olvidar nunca añadir la coletilla fastidiosa de que los estados psicológicos son además, para el realismo intencional, causas de las acciones y, por lo tanto, causas de ciertos movimientos corporales que constituyen la realización de esas acciones. Aunque no haya leyes psicofísicas, sí que hay correlaciones causales. Por lo tanto, es importante saber en qué sentido se puede hablar de los sucesos psicológicos como "reglas del cuerpo", y como reglas que, además, tienen capacidad causal. Adviértase que la dificultad que se encierra en esta manera de hablar tiene su origen en que la analogía con las reglas del baloncesto y del ajedrez nos abandona en este punto. El problema es sencillamente que no sabemos cómo hablar de las reglas del baloncesto como causas de la serie de movimientos que han constituido enteramente la realización de un partido.

Este problema causal era, de hecho, uno de los que necesitaba ser tratado mucho más a fondo desde hacía tiempo. Hasta este instante, el foco de la atención se ha concentrado especialmente en la aclaración sistemática de cómo es posible sostener el dualismo de las propiedades físicas y mentales, pero el monismo materialista de las sustancias. Para alcanzar este fin, las argumentaciones habían rondado en torno a la posibilidad de

defender coherentemente la idea de la identidad contingente entre sucesos psicológicos y físicos, idea con la que parece estar especialmente comprometida aquella posición ontológica mentalista. En el camino de esta defensa, por lo demás tan transitado por propios y extraños, uno de los logros parciales importantes había sido la poca credibilidad que tiene en estos tiempos sostener el fisicalismo-tipo, esto es, la tesis de que los predicados psicológicos y los predicados físicos son nomológicamente coextensivos. Pero ahora hay que volver a encarar el problema peliagudo de hacer cuadrar esta línea de argumentación con la tesis--de la que parece no podemos desprendernos desde el punto de vista del realismo intencional--según la cual el suceso que consiste en que x satisfaga un predicado psicológico está causalmente correlacionado con el suceso que consiste en que x satisfaga un predicado físico. Porque todo es uno y el mismo embrollo interesantísimo. Los sucesos psicológicos son distintos de los sucesos físicos--en el sentido de satisfacer predicados distintos--, pero unos y otros son contingentemente idénticos: unos y otros son sucesos de la misma sustancia material. Se puede enunciar de otra manera la idea de identidad contingente entre sucesos psicológicos y físicos diciendo que todo suceso psicológico satisface una u otra descripción física (o neurofisiológica) en virtud de la cual cae bajo una ley física (o neurofisiológica), pero que no todo género psicológico, no todo predicado que aparece en la formulación de las leyes psicológicas, es un predicado nomológicamente coextensivo con un predicado físico o neurofisiológico, aunque siempre que se está en un estado psicológico se está en uno u otro estado físico. El aire contradictorio que tiene afirmar estas dos cosas--la distinción entre sucesos psicológicos y físicos, y, si embargo, su identidad contingente--, se purificaba defendiendo la tesis de que no hay descripciones físicas uniformes de muchos sucesos que caen bajo generalizaciones científicamente relevantes,

tesis que llevaba a defender la imposibilidad del fisicalismo-tipo y la necesidad del fisicalismo-muestra.

Como se puede apreciar a lo largo de toda la exposición de este párrafo, un gran número de discusiones en la filosofía de la mente contemporánea se puede sintetizar bajo el título de "pugna de las descripciones". Cómo se describa un suceso determinará bajo qué leyes caerá, si es que esa descripción resulta finalmente científicamente relevante, es decir, si es que los sucesos así descritos caen bajo géneros naturales y satisfacen, consecuentemente, los predicados que los expresan. A su vez, si ese suceso cae bajo leyes, es que ese suceso es causalmente relevante bajo esas descripciones. Por consiguiente, si hay generalizaciones científicamente relevantes para muchas especies distintas de sucesos, ello se debe, muy probablemente, a que esos sucesos tienen papeles causales específicos respecto de otros sucesos. Los sucesos que caen bajo generalizaciones científicamente relevantes tienen la función característica de servir como miembros en cadenas causales. Precisamente, por su función causal es como las ciencias tratan de individuar los sucesos propios de sus dominios. Desde este punto de vista, la polémica en este caso se origina a la hora de decidir si las descripciones psicológicas de ciertos estados del organismo son o no son científicamente relevantes, es decir, si los estados así descritos tienen poderes causales. "Un vocabulario es apropiado... si sus términos designan solamente aquellas entidades que son claramente capaces de figurar en las transacciones causales" (Fodor 1980b, 76; véase también Davidson 1985, 117).

El nuevo reto que se presiente con la introducción de esta idea es el siguiente: si, de acuerdo con lo que se acaba de decir, los sucesos que caen bajo generalizaciones psicológicas tienen papeles causales específicos respecto de los sucesos que caen bajo generalizaciones físicas o neurofisiológicas y, por otra parte, hay causación psicofísica, pero los sucesos psicológicos son contingentemente idénticos con los sucesos físicos,

entonces, ¿cómo es que se puede seguir manteniendo la disparidad causal de los sucesos psicológicos y de los sucesos físicos, o neurofisiológicos, a la par que su identidad contingente? Si hay alguna manera de resolver el rompecabezas que planteó Sócrates en *Fedón*, desde luego ésta tiene que pasar inevitablemente por la solución de la dificultad expuesta. Sólo así se podrá comprender un poco mejor de qué manera está conectada la causa inteligente de las cosas con "aquello sin lo cual la causa no podría nunca ser causa", siendo esto último, a su vez, causa.

(d) El problema del dualismo causal.

Cuando se comentaron en el capítulo anterior las consecuencias y dificultades que planteaba el problema de la inteligencia encarnada, resaltó entre todas ellas aquella que aludía a la necesidad de integrar en la explicación global de las condiciones reales que hacen posible que los organismos realicen una determinada acción—especialmente si ésta implica movimientos corporales—todo aquello que la neurología tuviera que decir acerca de la etiología de los movimientos que constituyen esa acción. Sólo quiero recordar al lector que la dificultad principal consistía en que si la etiología fisiológica era suficiente para la explicación de los movimientos que constituyen la realización de una acción determinada, entonces la etiología psicológica sería superflua. Antes de abordar de nuevo esta cuestión con más precisión técnica y ver el modo en el que se podría solucionar, resultará de mucha utilidad encontrar algún ejemplo un poco más intuitivo que haga ver de qué manera podría ser tratado este problema. El ejemplo que voy a presentar para ilustrar esta situación es, una vez más, de Jaegwon Kim (1979, 37-38). Bajo el supuesto del realismo científico, existe la tentación de identificar el suceso que consiste en que este volumen de agua tenga esta temperatura (y que así describo), con el

suceso que consiste en que las moléculas de H₂O de este mismo volumen de agua tengan determinada cantidad media de energía cinética. La cuestión es saber si realmente debemos ceder a la tentación de identificar un suceso con el otro.

En un principio se podría contestar que esos dos sucesos son completamente diferentes, porque ambos satisfacen propiedades igualmente diferentes, aunque ambos sucesos coexistan espacio-temporalmente en la misma materia subyacente. Sucedería aquí quizás algo parejo a lo que ocurría en el caso trillado del color y la extensión de los objetos físicos. También el color y la extensión particulares de este objeto pueden coexistir espacial y temporalmente, pero no por ello afirmamos que tener color y tener extensión son la misma propiedad.

Sin embargo, en una segunda reflexión, sí que hay un argumento que induce a la identificación total de ambos sucesos. Este argumento hace uso del procedimiento, totalmente normal en las ciencias, de clasificar y, por lo tanto, de individuar los sucesos por sus poderes causales. En este sentido, se clasifica un suceso por sus poderes causales cuando no se cuenta ninguna propiedad de ese suceso como taxonómicamente relevante "a menos que afecte a sus poderes causales" (Fodor 1987, 38). Si un trozo de cera cae en este volumen de agua y se derrite, ¿cuál es la causa de que se derrita? Si se afirma, como es de esperar en un principio, que su causa es la alta temperatura del agua, entonces ¿cuál es el papel causal de la elevada energía cinética media de las moléculas del agua? Si se sostiene, por el contrario, que la causa de ese fenómeno es solamente este último suceso, entonces se plantea la cuestión inversa: la de asignar un papel causal a la elevada temperatura del agua. Es preferible, por lo tanto, identificar ambas propiedades y ambos sucesos para evitar el problema con el que nos encontraríamos si postulásemos dos causas diferentes, pero suficientes cada una de por sí, para la producción del mismo efecto. Si no realizamos esta

identificación, nos encontraremos con el típico problema de la sobredeterminación causal, completamente banal, de un mismo efecto. Este es el problema de la "causación dual" (Kim 1979, 36). La temperatura del agua y la alta energía cinética media de las moléculas, son causas duales de que la cera se derrita, si es que nos resistimos a identificar ambos sucesos.

Una manera posible y, por lo demás, muy socorrida, de solucionar este problema sería hablando de niveles diferentes de descripción para los mismos sucesos. Así es como, por ejemplo, propone proceder Searle (1983, 266). Cuando describimos este volumen de agua diciendo que tiene una temperatura muy alta, estamos realmente describiendo la elevada energía cinética de sus moléculas, sólo que en un nivel de descripción superior. Sobre la analogía de esta clase de ejemplos, la tentación de considerar las descripciones psicológicas de ciertos sucesos que tienen lugar en el cuerpo del organismo como descripciones que se realizan en un nivel de abstracción superior a aquel en el que se realizan las descripciones neurofisiológicas de esos mismos sucesos, es constante⁶. En definitiva, la tentación permanente cuando se llega a plantear así los problemas de las relaciones psicofísicas, es considerar los sucesos psicológicos como propiedades fenomenológicas, como "cualidades secundarias", aunque quizás no por ello menos reales, de las propiedades neurofisiológicas de cierta materia subyacente. Searle, por ejemplo, está dispuesto a hacer esta ampliación de la analogía para entender las relaciones lógicas entre los estados psicológicos y los estados del cerebro (1983, 267). De momento sólo quiero dejar anotado aquí un problema que trataré con más detalle posteriormente. ¿En qué sentido se puede decir que mis estados psicológicos

⁶ Como se recordará, la instalación en un nivel de descripción abstracto era uno de los rasgos fundamentales que ha adoptado la estrategia de investigación de la "naturaleza" de los pensamientos desde el punto de vista natural. Cf. el parágrafo 2 de este capítulo. Ahora se comprende un poco mejor el fundamento que tiene esta manera de proceder y la clase de argumentos sobre los que se apoya.

son cualidades secundarias? ¿No son las cualidades secundarias solamente cualidades para un sujeto?

Una vez que se establece esta teoría mínima acerca de los niveles de descripción de los sucesos que tienen lugar en una misma sustancia, cabe preguntarse si los sucesos que se describen utilizando el vocabulario que pertenece a cada uno de estos niveles se pueden considerar como causalmente activos bajo cada una de estas diferentes descripciones. Los sucesos neurofisiológicos que explican causalmente el suceso que se describe como "el cuerpo de Sócrates en esta posición" son, desde luego, causalmente activos y enteramente relevantes, pero, ¿lo son también los sucesos psicológicos que explican causalmente que Sócrates esté sentado en la cárcel?

Muchísimos autores--Mackie, Kim y Searle, por mencionar sólo a unos cuantos--han hecho ver el callejón sin salida en el que parecemos estar acorralados cuando planteamos así las cosas. Si se sostiene a la vez (i) el monismo materialista de las sustancias, (ii) el dualismo de propiedades, (iii) la idea de que las creencias y los deseos son causas de acciones cuya realización está constituida por movimientos corporales físicos y (iv) que las propiedades físicas y neurofisiológicas del organismo son ellas mismas causas suficientes de esos mismos movimientos, entonces parece que hay que eliminar una de estas tesis. Pero, ¿cuál exactamente estaríamos dispuestos a sacrificar? Si eliminamos (iii) porque, por ejemplo, consideramos que (iv) es suficiente para explicar causalmente la conducta del organismo, entonces nos comprometemos automáticamente con el epifenomenalismo, con la tesis de que el ser psíquico es causalmente inerte, con la tesis de que los sucesos mentales no tienen ningún papel causal en la producción de la conducta. Fácilmente, además, nos comprometemos también con el materialismo eliminativista, con la defensa de la identidad mente-cerebro. Precisamente, un argumento fuerte para defender el materialismo eliminativista--la nueva

versión del fisicalismo positivista--, es que el origen de todas las dificultades estriba en sostener la distinción fatal entre sucesos físicos y sucesos psicológicos, distinción que era, como se recordará, el punto de partida a que nos veíamos obligados de entrada cuando nos negamos a considerar las leyes puente como enunciados estrictos de identidad entre clases de sucesos psicológicos y físicos. Sin embargo, la tesis central del materialismo eliminativista parece que va en contra de todas nuestras intuiciones y que nos deja enfrentados con la difícil explicación de cuál sería el lugar de nuestras experiencias psíquicas conscientes que hablan en favor de la existencia indubitable de estados psicológicos; así como con la explicación de por qué pensamos machaconamente que estos estados psicológicos son causalmente activos. Por otra parte, si mantenemos nuestras intuiciones usuales y defendemos (iii) a la par que la tesis de que los sucesos neurofisiológicos constituyen cadenas causales que hacen que por fin ocurra el movimiento corporal que constituye la realización efectiva de la acción, entonces son esas cadenas causales, y no los estados psíquicos, las causas de la conducta. La responsabilidad causal sigue estando del lado físico y nosotros seguimos teniendo problemas.

Mackie, a modo de aclaración, sostiene que lo que está en juego en esta encrucijada, en la que corremos el riesgo de ser incapaces de tomar decididamente ningún camino, es un problema ontológico y no meramente lingüístico, como en ocasiones se piensa.

A veces se ha argumentado que el lenguaje podría cambiar de tal forma que hablar acerca de procesos cerebrales y cosas semejantes reemplazará a hablar acerca de experiencias. Bien, supóngase que podría cambiar y que de hecho ha cambiado: esto no clarificaría nuestro problema. Parte de este problema es si este lenguaje reformado sería adecuado, si cubriría el mundo, no meramente en el sentido de contar y predecir todos los acontecimientos concretos que caen bajo alguna descripción, sino también en el sentido de no dejar sin describir algunas características perfectamente reales y (al

menos para nosotros) interesantes. Por otro lado, se ha argumentado más plausiblemente que cualesquiera que sean los descubrimientos a los que llegue la ciencia y cualesquiera que sean los argumentos que los filósofos fisicalistas desplieguen, siempre continuaremos hablando en términos de experiencias y de pensamiento intencional, y que hablar acerca de las acciones pertenece firmemente a esta rama del lenguaje ordinario. Pero incluso si es así, esto tampoco clarificaría el problema en el sentido opuesto: la cuestión permanece en el sentido de saber si esta forma persistente de hablar expresa una verdad o un error sistemático... Que el lenguaje de la conciencia sea prácticamente indispensable podría muy bien resultar, y ser, evidencia indirecta de la presencia de propiedades mentales irreducibles, pero no clarifica la cuestión directamente o por sí mismo. Nuestro problema es acerca de qué propiedades hay de hecho y de cuáles son sus relaciones causales; no es acerca de nuestro lenguaje actual y de nuestro esquema conceptual, ni es tampoco acerca de cómo éstos podrían o no podrían ser alterados. (Mackie 1979, 20).

Una vez que se han mostrado las dificultades que existen a la hora de repartir poderes causales a los sucesos psicológicos y a los sucesos neurofisiológicos y ver que aquí, antes que nada, se trata de un problema ontológico, Mackie llama la atención sobre el hecho de que todos estos problemas surgen cuando se asumen varias piezas importantes de la teoría general de la causalidad. Creo que haber sabido ver nítidamente estos supuestos es una de las contribuciones principales de Mackie al problema de la causalidad en el ámbito de la filosofía de la psicología.

La primera de las asunciones que señala Mackie es que el conjunto de los problemas presentes no hubiera tenido ningún sentido si no se asume que los sucesos que tienen lugar en ciertos objetos, cuando esos sucesos satisfacen ciertas propiedades, son causalmente relevantes y que, por lo tanto, no hay que considerar exclusivamente como causalmente relevante al objeto concreto tomado como un todo. Si fuera el organismo como un todo el elemento mínimo que hubiera de ser considerado como causalmente operativo para la producción de la conducta, entonces no tendría ningún sentido preguntarse por la contribución causal que los distintos sucesos que

se dan en él tienen para la producción de esa conducta. No habría ningún suceso que fuera causalmente inerte ya que lo que es causalmente operativo es el todo orgánico concreto (Mackie 1979, 22).

El segundo elemento importante de la teoría de la causación que es relevante para la discusión del dualismo causal, es el problema que Mackie rotula como "dirección de la causación" (1979, 22). Es solamente bajo el supuesto implícito de que toda relación causal particular es asimétrica cuando se puede preguntar con sentido si podría haber un sendero causal desde lo psicológico a lo físico que fuera independiente del sendero causal desde lo físico a lo físico. Solamente si el sendero causal de lo psicológico a lo físico no es independiente del otro sendero causal, es cuando tiene sentido plantear el problema del epifenomenalismo, dado que el sendero de lo físico a lo físico es el que lleva el peso de la relación causal, siendo el otro inerte y redundante.

La tercera asunción es que los sucesos psicológicos tienen que tener la misma duración temporal que sus bases neurofisiológicas, de tal manera que si hay asimetría causal entre lo neurofisiológico y lo psicológico, ésta debe consistir en algo más que la simple sucesión en el tiempo. La co-temporalidad está implícita en su identidad (contingente).

Mackie cifra de ahora en adelante la llave del problema del dualismo causal en entender adecuadamente en qué consiste la asimetría causal (1979, 22). Un ejemplo de von Wright le sirve para comenzar la exploración. Pensemos en una caja con dos botones en su parte superior. Los botones—llamémosles "A" y "B"—están conectados de tal manera en el interior de la caja que siempre que se presiona uno de ellos los dos bajan a la vez. En estas circunstancias, el que *A* baje es una condición suficiente y necesaria para que baje también *B*, y viceversa. Sin embargo, Mackie defiende que el que *A* baje cuando lo presiono con el dedo es un suceso causalmente anterior al hecho de que *B* baje, aunque tanto *A* como *B*

bajen al mismo tiempo cuando presiono *A*. Lo que se pone de manifiesto con este ejemplo es que nuestro concepto de prioridad causal—que *A* baje es un suceso que es causalmente previo a que lo haga *B*—, es diferente de nuestro concepto de anterioridad temporal—*A* y *B* bajan a la vez, si se asume la idealización de la rigidez absoluta de los materiales. De esta manera, el concepto de anterioridad causal no descansa necesariamente en el de anterioridad temporal, sino sobre todo, según Mackie, en el de la continuidad de la causación. El movimiento descendente del botón *A* es causalmente previo al movimiento descendente del botón *B*, porque mi dedo ha presionado de hecho *A* y no *B*, y es razonable pensar que la presión de mi dedo sobre el botón *A* es causalmente previa tanto al movimiento descendente de *A* como de *B*. En este caso se dice que el movimiento de *A* está en continuidad causal entre el movimiento de presión de mi dedo y el movimiento descendente del botón *B*. Sin embargo, Mackie sostiene que el concepto de prioridad causal no descansa en el concepto de continuidad causal, porque el mismo problema que se planteaba a propósito de la prioridad causal del suceso que consiste en que *A* baje respecto del suceso que consiste en que *B* baje, se plantearía también respecto de la prioridad causal del movimiento de presión de mi dedo o de mi intención de presionar *A* respecto del suceso que consiste en que el botón *A* descienda. Más bien, lo que se quiere afirmar cuando se dice que puede haber dos sucesos simultáneos, y que cada uno de ellos es razón suficiente y necesaria del otro, pero que uno es causalmente previo al otro, es que esa relación de simultaneidad podría ser diferente en otras circunstancias. Si se rompiera la conexión que une a *A* con *B* en el interior de la caja, entonces el movimiento de un botón no sería una condición necesaria y suficiente para el movimiento del otro. Este hecho condicional es el que lleva a afirmar que el movimiento del botón *B* está condicionado causalmente respecto del movimiento del botón *A* y que, por lo tanto, el

movimiento de *A* es causalmente previo al movimiento de *B*, aunque ambos sean simultáneos. Son, por tanto, ciertos experimentos que buscan variaciones funcionales los que sugieren y confirman la dirección de la causalidad. Ejemplos semejantes a éste, junto con otros un poco más complicados, permiten a Mackie establecer una condición general para la prioridad causal y condicional de dos sucesos.

Si en una ocasión particular el que *A* haga *X* está causalmente relacionado con que *B* haga *Y*, y si ellos no hubieran estado relacionados de esa manera, sino que las cosas hubieran sido de una manera tan diferente como fuera posible, pero *A* todavía hubiera hecho *X*, aunque *B* no hubiera (o pudiera no haber) hecho *Y*, entonces el que *A* haga *X* es condicional y causalmente previo a que *B* haga *Y*. (Mackie 1979, 24).

Con este valioso principio en la mano, Mackie se enfrenta a su aplicación al caso de la causación psicofísica. La dificultad radicaba, recuérdese, en que si el sendero causal neurofisiológico era suficiente para la producción del conjunto de movimientos que constituyen la realización de una determinada acción, entonces la causación mental sería fútil, si no inexistente. Este principio permite formular exactamente en qué consistiría la inutilidad de la causación mental.

Si la conexión entre las bases y los sucesos neurofisiológicos y los sucesos psicológicos hubiera estado ausente, permaneciendo el resto de las condiciones iguales, entonces los sucesos neurofisiológicos hubieran sido exactamente los mismos, aunque no hubieran estado presentes los sucesos psicológicos. (Mackie 1979, 26).

La consecuencia condicional que extrae Mackie de este segundo principio menor es que si nunca puede faltar la conexión entre los sucesos neurofisiológicos y los sucesos psicológicos que se dan en los organismos, entonces no hay dirección de condicionalidad entre ellos y ninguno es

causalmente prioritario respecto del otro (1979, 26). En otras palabras, dos sucesos, ninguno de los cuales podría haber ocurrido sin el otro, no compiten por el papel causal: aun cuando uno de ellos sea una causa suficiente al igual que necesaria de algún efecto, el otro será también una causa suficiente y necesaria de ese mismo efecto (Mackie 1979, 27). Siempre que un suceso cause, causa también el suceso inseparable.

Jaegwon Kim ha reformulado esta consecuencia de una manera ligeramente diferente bajo el rótulo de "relación determinativa" (*determinative relation*): si dos sucesos son tales que su conexión no puede faltar, entonces no pueden ser felizmente considerados respectivamente como causa y efecto (1979, 45)⁷.

Como se ve, en cada esquina del laberinto las opciones que surgen son múltiples. Como éste es un auténtico laberinto, cualquiera de las opciones parece ser tan problemática como las otras. La siguiente esquina, cualquiera que sea el camino que se haya tomado en la esquina anterior, espera con nuevas complicaciones. En este momento quiero hacer una observación para destacar el grado de complejidad que se puede alcanzar. La solución de Mackie sería muy elegante y coherente con el hilo argumental precedente. Si los sucesos psicológicos y físicos son contingentemente idénticos y si uno es causa suficiente y necesaria de un efecto, entonces el otro también lo es. La formulación de Kim es más problemática, pero igualmente plausible. Si un suceso psicológico es contingentemente idéntico con uno físico, entonces aquél no puede ser considerado como causa de éste. Lo que es idéntico, aunque sea "contingentemente", no puede ser causa de sí mismo. Para eso hace falta ser Dios.

⁷ Este es un punto importante en el que se separa Searle de las tesis de Mackie y de Kim. Para Searle es claro que la relación entre un suceso y el otro sí que sería de causación. Véase (Searle 1983, 265-266).

(e) Superveniencia.

Jaegwon Kim, en una serie persuasiva de artículos⁸, ha puesto de relieve de qué manera se podrían conjugar armoniosamente gran parte de las tesis que han ido apareciendo a lo largo de la discusión de este párrafo e introduce algunos refinamientos conceptuales muy importantes y esclarecedores. El primero y principal entre éstos se refiere a lo confuso que puede llegar a resultar distinguir entre clases de identidades; entre identidades (contingentes) de sucesos (token-identity) e identidades de tipos (type-identity). En lugar de hablar de identidades-muestra e identidades-tipo entre lo psicológico y lo físico, propone hablar de la "superveniencia" de unos sucesos en los otros (Kim 1979, 44).

Pero, ¿qué matiz de claridad o, incluso, de novedad conceptual introduce este importantísimo concepto de "superveniencia" que tan ampliamente ha sido utilizado para "solucionar" el problema mente-cuerpo? Como se sabe, el concepto de "superveniencia", también llamado "propiedad consecucional", fue introducido y explotado ampliamente por los filósofos morales británicos G. E. Moore y R. H. Hare. Con él, estos filósofos quisieron resaltar que dos sucesos o estados de cosas no pueden concordar en todas sus características descriptivas naturalistas y diferir, en cambio, en sus características evaluativas o morales, aun cuando las propiedades evaluativas y morales no se puedan definir naturalistamente.

Sobre este modelo, la definición más amplia del concepto de superveniencia que ofrece Kim dice lo siguiente:

⁸ Jaegwon Kim. 1978. "Supervenience and Nomological Incommensurables." American Philosophical Quarterly 15: 149-156; 1979. "Causality, Identity, and Supervenience in the Mind-Body Problem." Midwest Studies in Philosophy. Studies in Metaphysics. Eds. Peter A. French et al. Minneapolis: University of Minnesota Press: 31-49; 1982. "Psychophysical Supervenience." Philosophical Studies 41: 51-70.

Sea M <una familia de propiedades que> superviene en <una familia de propiedades> N , y sea P una propiedad de M tal que un objeto x tiene P . Entonces un subconjunto B de N es *una propiedad base de superveniencia* ("base de superveniencia", para abreviar) del suceso que consiste en que x tenga P justo en el caso de que B sea un conjunto mínimo tal que x tiene las propiedades en B , y cualquier otra cosa que tenga estas propiedades tiene P . (1979, 43).

Esta definición de superveniencia funciona como una especie de compilación general en relación con la amplia gama de situaciones ontológicamente enrevesadas que se ha estado considerando más arriba. En efecto, la identidad contingente entre muestras de estados psicológicos y muestras de estados físicos se resuelve, gracias a este concepto, de la siguiente manera. La familia de las propiedades psicológicas superviene en la familia de las propiedades físicas, si hay un organismo que tiene una de las propiedades P de la familia de propiedades psicológicas—por ejemplo, está en un estado determinado de creencia—y es el caso que satisface un subconjunto de propiedades físicas F tal que estas propiedades serán la base de la superveniencia del suceso que consiste en que ese organismo satisfaga P , cuando el subconjunto de propiedades F sea un conjunto mínimo tal que el organismo tiene F y cualquier otra cosa que tenga ese subconjunto de propiedades tiene también P .

La idea sobre la que se quiere llamar la atención cuando se habla de conjuntos mínimos en la definición de la superveniencia, es que esos conjuntos deben de ser suficientes para que se dé la propiedad psicológica superveniente. Es altamente probable que las propiedades físicas de las piedras no sean un conjunto mínimo que sirva para la superveniencia de propiedades psicológicas.

Por otra parte, dos subconjuntos diferentes del conjunto de las propiedades físicas pueden ser la base de superveniencia del mismo suceso psicológico. Como ya se estableció después de argumentar en paralelo con

el caso de la economía, diferentes estados físicos pueden servir de base de superveniencia al mismo estado psicológico. Así, pues, el conjunto de las propiedades que supervienen y el conjunto de las propiedades de superveniencia no son nomológicamente coextensivas. Esta es, exactamente, una de las consecuencias inmediatas de la superveniencia de propiedades: que aunque haya correlaciones entre la familia de propiedades que superviene y la familia de propiedades sobre las que superviene aquélla, estas correlaciones no son nomológicas. Pero que haya correlaciones indica que entre las propiedades que son la base de la superveniencia *hay* una propiedad o conjunto mínimo de propiedades que es una razón suficiente para que se dé la propiedad que superviene.

Se sigue, por consiguiente, que si las propiedades psicológicas supervienen sobre las propiedades físicas, entonces toda instanciación de una propiedad psicológica tendrá una razón suficiente en las propiedades físicas" (Jaegwon 1979, 42; véase también Fodor 1975, 20-22)⁹.

En lo que se refiere al problema de la causación psicofísica, dice Kim que la solución que él propone "está basada en la idea de que el papel causal de un suceso, especialmente de un suceso nomológicamente inconmensurable, se puede explicar por medio del papel causal de un suceso

⁹ Jaegwon afirma que no hay razones para pensar que no haya leyes psicofísicas, simplemente porque no haya correlaciones legales entre tipos de estados psicológicos y tipos de estados físicos. Si hay correlaciones, afirmación que nadie pone razonablemente en duda, no hay motivo para pensar que no sean "legales" (lawlike). "Ellas parecen ser la clase de correlaciones empíricas que son objeto de confirmación por observación de instancias favorables y que pueden soportar contrafactuales" (Kim 1979, 34). Para Kim, si hay superveniencia de las propiedades psicológicas sobre las propiedades físicas, entonces es que tiene que haber correlaciones psicofísicas legales. "Se puede demostrar que para cada propiedad P en la familia superviniente M que esté instanciada en el dominio D, existe en el conjunto base N una propiedad Q que es una condición suficiente para ella....Esta correlación afirma que cualquier organismo o estructura que ejemplifique la propiedad física Q también ejemplificará la propiedad psicológica P" (Kim 1979, 42). Justo la tesis contraria es la de Davidson y Fodor.

sobre el que superviene" (1979, 49). Supongamos que P y Q son dos sucesos psicológicos que supervienen, respectivamente, a los sucesos físicos P^* y Q^* . Si hay relaciones causales entre P y Q , entonces esas relaciones ocurren a lo largo de un sendero causal que incluye necesariamente y que no puede ser independiente del sendero causal que une los sucesos físicos P^* y Q^* (Kim 1979, 46-47). Es de esta manera como se evita el problema de la dualidad causal. El concepto de dualidad causal estaba en estrecha relación con el de la existencia de dos sendas causales independientes pero suficientes cada una de por sí para la producción del mismo efecto. Ahora se puede evitar este problema. Simplemente no hay posibilidad alguna de que un suceso que satisface una propiedad del conjunto de propiedades que supervienen, pueda causar a lo largo de un sendero causal que no incluya esencialmente el sendero causal de los sucesos de la familia que constituye la base de la superveniencia.

(f) **Resumen.**

A la luz de todas estas reflexiones se puede intentar sintetizar y formular lo que llamaré "los seis dogmas del ser psíquico". Ellos resumen de una manera aproximada cuál es la imagen que parece subyacer a un sinnúmero de teorías de filosofía de la mente que se encaran con el problema de las difíciles relaciones mente-cuerpo y el reduccionismo fisicalista estricto. Son el plano esquemático de las esquinas del laberinto. Aquí están los seis dogmas del ser psíquico, cuya independencia, como puede apreciarse, es muy relativa:

1º. *El fisicalismo-muestra es verdadero.* Todo suceso psicológico individual tiene una descripción verdadera en el vocabulario de la física en virtud de la cual cae bajo leyes físicas. A veces se dice, quizás confusamente, que todo suceso psicológico individual es contingentemente

idéntico con un suceso físico individual. Decir que el fisicalismo-muestra es verdadero es lo mismo que defender el monismo materialista a propósito de las sustancias reales.

2º. *El fisicalismo-tipo es falso.* Los predicados psicológicos y los predicados físicos no son nomológicamente coextensivos. Como ya se vio, esta tesis afirma que en los organismos se dan sucesos que satisfacen tanto predicados psicológicos como predicados físicos, aunque unos y otros no sean nomológicamente coextensivos.

3º. *Es preciso defender la superveniencia.* No puede suceder--en un sentido de necesidad todavía por especificar--que dos organismos estén en estados psicológicos diferentes y que no difieran en ninguno de sus estados físicos. O, lo que es lo mismo, si dos organismos son absolutamente indistinguibles desde el punto de vista físico (o microfísico), entonces tendrán en común toda su vida psicológica. En este sentido se dice que los estados psicológicos de un organismo supervienen en sus estados neurofisiológicos, si y sólo si sus estados neurofisiológicos difieren siempre que difieran sus estados psicológicos.

4º. *El psicofuncionalismo es verdadero.* Aunque dos organismos difieran en sus estados físicos no diferirán por ello necesariamente en sus estados psicológicos.

5º. Desde el punto de vista de la psicología popular, ciertos sucesos psicológicos, como las creencias y los deseos, se consideran *causas* de todas aquellas acciones que, al menos, implican movimientos corporales.

6º. Las propiedades físicas y neurofisiológicas de los sucesos psicológicos individuales son *causas suficientes* de esos movimientos corporales.

Según estos seis dogmas, el ser psíquico introduciría una especie de desorden peculiar en el mundo. Es el anomalismo de lo mental. Lo que resulta extraño es defender a la vez 1, 2 y 5. Pero, ¿qué es lo

verdaderamente anómalo? Veamos si estos seis dogmas sirven para otras propiedades y objetos naturales. El fisicalismo-muestra es verdadero, por ejemplo, de los sucesos geológicos. Cualquier suceso geológico individual--ser *esta* montaña--tiene una descripción verdadera en el vocabulario de la física en virtud de la cual cae bajo leyes físicas. El fisicalismo-tipo es falso para la geología. Los predicados geológicos (ser una montaña) y los predicados físicos no son seguramente nomológicamente coextensivos. La superveniencia es verdadera para la geología. Cuando dos montañas difieran en sus estados geológicos, diferirán en sus estados físicos. El "geofuncionalismo", en cambio, *es falso*. Cuando dos montañas difieran en sus estados físicos, *diferirán también* en sus estados geológicos. Ser más alta u ocupar más superficie o ser más fácil de erosionar, son todos ellos, a buen recaudo, estados y propiedades geológicamente relevantes.

Cuando la discusión llega a este punto, parece que se puede presentar un cúmulo de argumentos y conceptos que es común para todos aquellos que no pueden ser hasta el final, por considerarlas posiciones teóricas insostenibles, ni fisicalistas positivistas estrictos, que defienden el monismo de las sustancias y el monismo de las propiedades, ni dualistas cartesianos, que defienden el dualismo de las sustancias y el dualismo de las propiedades. La búsqueda de un camino intermedio razonable entre unos y otros ha llevado a una u otra forma de "mentalismo", a defender ontológicamente el monismo de las sustancias, pero el dualismo de las propiedades físicas y psicológicas. Todo un sofisticado sistema de conceptos se puso en marcha para sostener esta posición ontológica en la que parece que hay que hacer verdaderas filigranas de equilibrio para mantenerse en ella. La identidad contingente entre sucesos físicos y psicológicos, pero, por otra parte, su diferencia; la superveniencia y el psicofuncionalismo, son algunas de las tesis claves que aparecen reiteradamente en cientos de discusiones y que abogan en favor de esta tesis intermedia. Si bien mucho

más complicada y con muchos más recovecos que la del fisicalismo positivista estricto o que la del cartesianismo, esta posición iluminaría mucho mejor la ontología materialista no reduccionista que se necesita para resolver el problema mente-cuerpo.

Sólo una nota última para avivar el recuerdo. A lo largo de este párrafo, la búsqueda estaba motivada fundamentalmente por el deseo de escapar de las férreas garras con que la teoría de la ciencia positivista apresó normativamente a la psicología cuando este campo del conocimiento humano quiso erigirse en verdadera ciencia objetiva. Las argumentaciones precedentes sólo han querido incidir persistentemente en la idea de que no hay razones de principio para sostener que los referentes de los enunciados teóricos de la psicología se refieran exclusivamente a los estados de cosas físicos del organismo. Pero, si no se refieren expresamente a estados físicos del organismo, ¿a qué se refieren exactamente? Una vez que se ha tomado la decisión de considerar los predicados psicológicos como predicados que tienen el mismo estatuto que los predicados de las otras ciencias especiales y como aquellos predicados que deben entrar a formar parte esencial de las teorías que son racionalmente pertinentes para explicar cierta clase de correlaciones funcionales empíricas, haciendo, por tanto, de los estados mentales *entidades teóricas*, queda la tarea de perfilar mejor qué clase de entidades y de estados de cosas son los referentes de estos predicados psicológicos, cuáles son sus propiedades y de qué manera podrían cumplir su función causal sin ser confundidas con los estados físicos del organismo. Llevar a realización adecuada este propósito constituye el objetivo de los

siguientes párrafos.

& 6. La situación experimental.

La psicología quiere fundamentar sus tesis y teorías sobre experimentos empíricos, quiere contrastarlas y discutir las con la realidad, en la medida en que esto sea posible en la ciencia empírica. El psicólogo formula, así, hipótesis y diseña experimentos más o menos complejos con las miras teóricas puestas en la verificación de esas hipótesis. En estas características, la psicología cognitiva no pretende apartarse tampoco de la manera de proceder propia de otras ciencias. Sin embargo, creo que existen algunos rasgos genéricos muy curiosos que comparten las situaciones experimentales en psicología cognitiva y que hacen que esas situaciones experimentales no se parezcan mucho a las que se dan en otras ramas de la ciencia. Por este motivo, creo que es de sumo interés describir genéricamente estos rasgos esenciales. Esta tarea, ciertamente no muy usual en la reflexión sobre los fundamentos de la psicología cognitiva, tiene, a pesar de todo, una importancia extrema si es que hay verdadero interés por acercarse con rigor a estos problemas radicales de fundamentación. Además, tenerlos presentes ayudará a comprender la manera de proceder posterior en este capítulo y permitirá la evaluación crítica de los resultados teóricos que se obtengan.

La primera descripción universalísima de las situaciones experimentales en psicología cognitiva, pero que, sin embargo, va directamente al corazón de sus características más importantes, tiene mucho que ver con la manera en que el psicólogo se enfrenta a su objeto de estudio. Desde este punto de vista, me interesa destacar que el teórico que ha diseñado la situación experimental "sabe mucho más" acerca de ella que el sujeto de experimentación, tanto desde el punto de vista teórico, como desde el

práctico; sabe más que éste acerca de cómo es el "mundo" en ese preciso momento. Es el punto de vista de la tercera persona, heterofenomenológico, completamente necesario para el diseño mismo del experimento psicológico. Para enunciarlo de una manera clásica: una de las condiciones de posibilidad para que se puedan contrastar las hipótesis teóricas de la ciencia psicológica cognitiva es la adopción por parte del psicólogo, desde el inicio mismo, de este punto de vista heterofenomenológico respecto del sujeto de experimentación. La investigación empírica de la intencionalidad, de ser posible, tendría que partir necesariamente de esta clase de situaciones. Aunque el contenido de estas afirmaciones quizás sólo se entienda ahora de una manera incompleta, se hará mucho más exacto y palpable en cuanto se analicen algunos experimentos típicos de psicología cognitiva y se vea realmente, sobre su base, el sentido y alcance de este primer rasgo descriptivo.

Por otra parte, en la situación heterofenomenológica el psicólogo cuenta de antemano con que el sujeto experimental que tiene enfrente de sí es un ser psicológicamente "semejante" a él, al menos en el respecto que está investigando. Esta sería la segunda característica descriptiva fundamental de las situaciones experimentales. Si se tiene en cuenta esta presuposición, la rata en el laberinto, o un ingenio mecánico con el que se juega al ajedrez, son tan "semejantes" al psicólogo como lo pueda ser cualquier otro humano. La idea insuprimible de "semejanza psicológica" desvela, además, un equívoco que ha cristalizado en una forma de expresión típica y que, además, parece definir a la psicología de este siglo: la de considerar al sujeto de investigación como "una caja negra". En realidad, sin embargo, nunca ha existido nada parecido a una "caja negra" como objeto de la psicología, algo acerca de lo cual no se supiera absolutamente nada desde el punto de vista psicológico. La caja negra es un puro mito. Siempre que el psicólogo hace un experimento psicológico para fundamentar

empíricamente una teoría conoce y presupone muchas cosas respecto del tipo de ser psicológico del organismo que es sujeto de experimentación. "Habla como hablaría yo, percibe como percibiría yo, se comporta como me comportaría yo, hace lo que haría yo", se dice calladamente el psicólogo.

La tercera propiedad descriptiva de la situación experimental ha sido mencionada ya extensamente en otros varios lugares y aquí sólo voy a enunciarla brevemente. Esta consiste en que el psicólogo cuenta con la evidencia práctica y empírica preexistente de que existen correlaciones y conexiones entre el comportamiento público observable del sujeto de experimentación y los procesos y estados mentales, inobservables por esencia, que tienen lugar en él.

Estos tres rasgos descriptivos--el abismo epistemológico y práctico insuprimible que existe entre el psicólogo que teoriza y experimenta y el sujeto de experimentación, la idea de semejanza psicológica entre uno y otro y las correlaciones supuestas entre el comportamiento del organismo y los procesos y estados mentales--, son condiciones absolutamente necesarias para que se puedan construir hipótesis y modelos científicos sobre el funcionamiento mental real de los sujetos y para que se puedan poner a prueba dichas hipótesis psicológicas. Son condiciones para que se pueda hacer ciencia psicológica. La relación entre esas tres características de la situación experimental se puede expresar sucintamente de la siguiente manera. Entre lo que sabe sobre la situación quien diseña el experimento y lo que dice sobre ella o hace en ella el sujeto con su conducta, se introduce la cuña de los constructos postulacionales de la psicología científica. Esta cuña es la que ha de explicar las correlaciones entre las conductas observadas y los estados y procesos mentales del organismo. El psicólogo, sin embargo, ya sabe que sus experimentos versan sobre la percepción, sobre la imaginación, sobre la memoria, la atención o el

razonamiento. Por consiguiente, él tiene que saber ya de alguna manera, quizás superficial e imperfecta, qué es percibir, imaginar, recordar, atender o razonar, antes de elaborar experimentos sobre estos fenómenos psicológicos. El psicólogo supone sin más que el sujeto de experimentación es semejante a él en relación a estos fenómenos psicológicos que él quiere investigar empíricamente. Por otra parte, si los constructos hipotéticos del psicólogo son verdaderos, habrán de ser generalizables y aplicables al propio teórico de la psicología, en virtud de la semejanza psicológica entre éste y el sujeto de experimentación. Estos constructos postulacionales, como ya se ha insinuado en otras ocasiones, son los que nos comprometen ontológicamente con entidades "mentales" que poseen unas propiedades antes insospechadas cuando nos manteníamos enteramente en el nivel de descripción fenomenológica de los estados mentales. Gracias a la aplicación de los métodos hipotéticos al objeto de estudio de la psicología--la explicación causal del comportamiento en un sentido amplísimo del término "comportamiento"--la psicología se alza ante nuestros ojos como una de las ciencias especiales. Además, no son pocos los filósofos de la mente que creen firmemente que estas entidades mentales postuladas, dado el camino metodológico racional que nos ha conducido hasta ellas, son las verdaderas entidades mentales de las que hablamos inadvertidamente cuando, instalados en el suelo básico de la psicología popular, sostenemos que las actitudes proposicionales tienen a la vez propiedades semánticas y causales. De lo que se trata cuando se diseñan experimentos psicológicos es de "pillar" a la mente y, en un cierto sentido, a nosotros mismos, en un comportamiento extraño, curioso, con el que no se contaba enteramente de antemano y que nos llama la atención. Sobre lo que no entendemos es sobre lo que nos preguntamos racionalmente y sobre lo que queremos tener buenas teorías.

Es ahora, a la luz de estos tres rasgos descriptivos esenciales de la situación experimental, cuando se entiende mucho mejor la afirmación que

hacia en el primer capítulo, según la cual la psicología cognitiva no es plenamente ni psicología descriptiva o fenomenológica, ni psicología genética. Brentano, a quien se debe precisamente la división de las investigaciones psicológicas en estas dos grandes áreas--psicología descriptiva, o psicognosia (*Psychognosie*), o psicología pura, y psicología genética--, caracteriza la psicología descriptiva como aquella que estaría preocupada fundamentalmente por el análisis y catalogación de los distintos elementos y partes de la conciencia, así como por sus modos de unión esenciales. Por el contrario, frente al objeto de la psicología descriptiva, se alza el objeto de la psicología genética, esto es, el estudio de las leyes y condiciones--seguramente fisiológicas--que gobiernan la aparición y desaparición de estas Erscheinungen o elementos de la conciencia, condiciones con las que están "unidos originariamente" aquellos elementos (Brentano 1982, 1). Otra forma interesante que tiene Brentano de distinguir la psicognosia de la psicología genética o fisiológica, es constatando que la primera no se ocupa en absoluto de las *causas* que hacen que un determinado elemento de la conciencia aparezca, desaparezca o permanezca durante un tiempo ante "la mirada" del yo. La psicognosia se ocupa tan sólo de aquello de lo que hay conciencia interna. En contraposición, la psicología fisiológica tiene por objeto el estudio de las "precondiciones" que, en tanto que procesos físicos y químicos que tienen lugar en el cuerpo, son imprescindibles para que aparezca la conciencia. Esta doble contraposición entre la psicognosia y la psicología fisiológica da origen a otras dos nuevas aún más importantes si cabe. La distinción entre la primera clase de psicología y la segunda viene delimitada a su vez por aquello acerca de lo cual cabe intuición y por aquello acerca de lo que no cabe intuición respectivamente. Y todavía más. La psicognosia, al hacerse cargo de aquello de lo que cabe intuición o conciencia interna, tiene principios que son exactos y precisos. La psicología

fisiológica, por su parte, sólo puede aspirar a formular leyes imprecisas, leyes acerca del devenir, del llegar a ser de la conciencia.

Pero el psicólogo cognitivo no hace psicología descriptiva, porque en ningún caso demora sus investigaciones dando vueltas incesantemente alrededor de sus vivencias conscientes para intentar describirlas todo lo minuciosamente que pueda en sus partes genéricas y específicas. Sin embargo, como se mencionó en un párrafo anterior, la psicología cognitiva no desdeña en modo alguno la contribución del sujeto consciente. Al contrario, la considera muy valiosa. El apoyo en el comportamiento manifiesto, tanto verbal como no verbal, no excluye que la psicología científica contemporánea utilice como una herramienta casi esencial en muchos ámbitos de investigación--el ejemplo más obvio sería el de la psicolingüística--los enunciados introspectivos que los sujetos utilizan para expresar el contenido de sus estados mentales en relación a las tareas experimentales propuestas. Consultar las partes que los manuales de psicología cognitiva al uso dedican a exponer los principios metodológicos de la investigación basta para persuadirnos de la verdad de este hecho. También la praxis experimental da pruebas abundantísimas de él. Esta afirmación se pone claramente de manifiesto, por ejemplo, cuando se lee la manera en que Finke describe algunas de las técnicas que se utilizan normalmente para determinar los niveles de equivalencia funcional entre las imágenes y las percepciones de los objetos.

Estas técnicas <técnicas de scaling> permiten comparar directa y precisamente la forma en que la gente juzga los objetos observados con la manera en que juzgan las imágenes mentales. Dado que los *tipos de juicios* que la gente hace cuando *observa* los objetos frecuentemente *corresponden* con la manera en que operan *ciertos tipos de mecanismos de procesamiento visual...*, estas técnicas permiten averiguar si los *tipos de juicios* que la gente hace cuando *imagina* objetos podrían también corresponder con la manera en que operan

ciertos tipos de mecanismos de procesamiento visual. (1980, 116) (Los subrayados son míos).

Como se puede apreciar a la luz de este ejemplo--y se podría hacer igualmente a la luz de otros semejantes--, el ejercicio de la introspección y de la reflexión por parte de los sujetos no se lleva más allá de que éstos expresen sus juicios acerca de percepciones y de imágenes, de que expresen si recuerdan o no una lista de palabras o de frases, de si perciben o no con la misma longitud dos líneas o de si perciben una anciana con una nariz muy grande o una joven atractiva con un abrigo de piel, de si han comprendido o no una frase determinada, o de que reconozcan si ha sido un hombre o una mujer quien la ha pronunciado, etc. Nunca se le pide al sujeto que describa fenomenológicamente--y menos aún con el detalle, minuciosidad y complejidad técnica con que la filosofía de Husserl ha conseguido hacerlo--sus vivencias intencionales conscientes (las *cogitationes*) y sus modificaciones atencionales, los distintos elementos genéricos y partes específicas que necesariamente las componen, y las relaciones peculiares que guardan entre ellas. Menos aún se le pide al sujeto que evalúe racionalmente el grado de evidencia con que se le dan todos estos elementos y partes. Nada de esto interesa esencialmente al psicólogo cognitivo. ¿Por qué no? Sencillamente, porque siguiendo esta vía de investigación está persuadido de que nunca llegaría a tener una idea cabal del funcionamiento real de la mente, que es, al fin y al cabo, lo que finalmente le interesa, y porque los aspectos fenomenológicos de las experiencias conscientes es lo que, en algunos casos, hay precisamente que explicar. No se trata, por consiguiente, de que el psicólogo cognitivo esté de acuerdo o no con el sujeto de experimentación sobre los fundamentos racionales y epistemológicos que tiene para afirmar lo que afirma, o de que polemice con él sobre el grado de asentimiento y evidencia que está dispuesto a otorgar a sus juicios. En general no se trata de que discuta

con él sobre estos asuntos, por otra parte tan cruciales en las teorías del conocimiento clásicas. La cuestión debatida aquí, más bien, es la de construir modelos operativos que especifiquen y expliquen los procesos "internos" que han llevado al sujeto a emitir finalmente ciertos juicios introspectivos¹.

Se supone, además, que estos procesos no son enteramente describibles como procesos neurofisiológicos, sino como procesos que manipulan información y que se realizan en determinados mecanismos físicos y neurofisiológicos. El párrafo anterior estuvo dedicado enteramente a poner de relieve la lucha apasionada que existe en el campo de la filosofía de la psicología a propósito del problema clásico del dualismo mente-cuerpo. La polémica, que quizás tenga ahora un aspecto mucho más técnico del que tuvo en el pasado, se centró en el problema de la reducción de la psicología a la física. Los diversos contendientes interesados en este problema siguen argumentando entre sí y da la sensación de que lo seguirán haciendo en un futuro próximo. Aunque no haya nada decidido, sí que se puede indicar que gran parte de la reflexión actual en la filosofía de la psicología está orientada a procurar un asiento cada vez más firme a la idea de que la psicología se ocupa de sucesos y estados que no se pueden reducir a sucesos neurofisiológicos o físicos. Por esta clase de consideraciones, la psicología cognitiva no se sentiría cómoda si se la tildase simplemente como psicología genética, en especial si se sostiene que

¹ Se puede detectar un escepticismo generalizado a propósito del interés de los frutos teóricos que para la psicología pueda dar la investigación puramente fenomenológica de los estados mentales conscientes. Fodor sostiene, por ejemplo, que la distinción entre lo que un organismo humano hace y lo que hace su sistema nervioso puede ser muy importante para algunos fines como el de atribuir responsabilidad moral. Pero que sea importante para esos fines no quiere decir que sea importante para los fines de la psicología cognitiva. "...La distinción ordinaria entre lo que un organismo hace, piensa y sueña, y lo que le sucede al sistema nervioso y lo que sucede en el sistema nervioso, no parece ser terriblemente importante" para la investigación psicológica (1975, 52-53).

las leyes que gobiernan la aparición y desaparición de las Erscheinungen son sólo leyes empíricas neurofisiológicas.

& 7. Aplicaciones empíricas del método.

Con estos primeros pasos panorámicos por la estrategia metodológica general que la psicología cognitiva propone para investigar, desde el punto de vista científico-natural, qué son los pensamientos y cuáles son sus propiedades, y con el fin de apuntalar más firmemente el edificio que se va vislumbrando poco a poco con estas consideraciones iniciales, lo mejor que puede hacerse es, sin duda, concretarlas con varios ejemplos. Voy a proponer el análisis de tres clásicos, elegidos casi arbitrariamente entre la ingente producción bibliográfica en psicología cognitiva. Por lo demás, estos ejemplos están tan trillados que, como dicen con frecuencia los autores anglosajones, quien conozca ya esta parte de la historia que salte al párrafo siguiente. En su elección he tratado de buscar simplemente aplicaciones relativamente sencillas del método esbozado en los párrafos anteriores, para que se vea lo más intuitivamente posible la clase de explicaciones causales que este método puede ofrecer. Su análisis será, por otra parte, la base sobre la que se destacarán los compromisos ontológicos de la psicología cognitiva que desde hace tanto tiempo están buscándose.

El primero de los ejemplos que se analiza es un caso típico de comportamiento verbal; el segundo, uno en el que se mezcla el comportamiento verbal y el no verbal; y el tercer ejemplo, un caso de comportamiento no verbal. Con ellos no se pretende otra cosa más que dar una muestra representativa que sea capaz de transmitir "el sabor" genérico de las investigaciones teorico-empíricas en psicología cognitiva.

(a) La representación de los significados.

En este primer ejemplo, tanto el teórico que diseña el experimento para comprobar la verdad de las hipótesis de la teoría psicológica, como el sujeto de experimentación son dos miembros de la misma comunidad de hablantes. Los psicólogos son, en este caso, Lackner y Garrett (1973). La teoría psicolingüística que guía el diseño del experimento dice que los hablantes oyentes de un lenguaje natural atribuyen inconscientemente ciertas descripciones estructurales lingüísticas, de carácter supuestamente universal, a las oraciones del lenguaje. Estas descripciones estructurales están jerarquizadas en diversos niveles. Los más básicos son, de inferior a superior, el nivel fonético, el nivel sintáctico y el nivel semántico. Es sólo cuando la oración emitida por el hablante alcanza este último nivel de representación en el sistema lingüístico del oyente cuando se produce en él la comprensión propia del significado de la oración. Desde el punto de vista psicolingüístico, los hablantes y oyentes de un lenguaje se consideran como mecanismos que son capaces de asociar significados o mensajes con ondas sonoras, y viceversa. Una teoría psicolingüística es una teoría que quiere explicar las condiciones necesarias para que se pueda llevar a cabo esta asociación y formular hipótesis acerca de los mecanismos que la realizan.

Para evitar equívocos, hay que anticipar que ésta no es una teoría directamente relevante para discutir cómo ciertos objetos físicos--las grafías o las voces--, pueden tener significado. Tampoco es una teoría directamente relevante para contestar a la pregunta "¿qué son los significados de las oraciones o de las palabras?". Sólo es una teoría que sostiene que hay bases empíricas para afirmar que en el sistema lingüístico de los individuos debe haber representación y procesamiento inconsciente del significado y de otras propiedades lingüísticas relevantes de las oraciones. Se dice que la oración

emitida habrá de recibir una representación de sus propiedades fonéticas, sintácticas y semánticas antes de que se "produzca" su comprensión propia. Por lo demás, la consecuencia ontológica más directa de la teoría, de ser ésta verdadera, dice que para que pueda haber representación de las propiedades lingüísticas de la oración es preciso que previamente, como condición suya, haya un sistema representacional en el que se pueda representar estos niveles. Estas descripciones estructurales abstractas, por lo demás, se supone que son *psicológicamente reales*, esto es, que no se puede dar cuenta de cómo es posible que se den las asociaciones entre ondas sonoras y significados si no se interpretan esas estructuras abstractas que postula la teoría psicolingüística como partes reales del sistema psíquico y cognitivo de los individuos.

En este experimento se les presentan a los sujetos distintas oraciones a través de unos auriculares estereofónicos. Los sujetos deben atender exclusivamente a la información semántica de las oraciones que se les transmite por uno de los dos auriculares y desatender sistemáticamente la información que se les presenta por el otro. Para asegurar el éxito de esta tarea se procede, además, a la interferencia o ensombrecimiento (shadowing) de lo que se dice por el canal que debe ser desatendido. El modo concreto de materializar esta interferencia es, por una parte, disminuyendo considerablemente el volumen de este canal en relación con el canal que debe atenderse y, por otra, variando ampliamente sus contenidos. Este modo de interferencia es tan efectivo que los sujetos no pueden decir en la mayoría de los casos qué clase de material acústico se emitía por el canal desatendido; si se trataba de música, de ruidos sin sentido, de palabras sueltas, o de oraciones con pleno sentido.

En el experimento hay dos grupos de sujetos: un grupo de control y otro grupo crítico dividido, a su vez, en dos subgrupos. En la situación experimental los sujetos del grupo crítico escuchan clara y distintamente

una oración de significado ambiguo por el auricular al que deben atender. Por el canal desatendido se emite otra oración que, en condiciones normales, permitiría eliminar la ambigüedad significativa de la oración atendida al dar una información contextual apropiada para ella. Sin embargo, debido a los fenómenos de interferencia y a la falta de atención que los sujetos prestan a este canal, dichos sujetos no están en condiciones aparentes de utilizar conscientemente el significado de esa oración desatendida para eliminar la ambigüedad significativa de la oración atendida. La oración atendida con significado ambiguo era: (i) "the spy put out the torch as our signal to attack" ("el espía apagó/sacó la antorcha como señal para que atacáramos"). El verbo "to put out", entre otros significados, tiene en inglés los de "apagar" y "sacar". En este hecho semántico reside la ambigüedad significativa de la oración en la que aparece dicho verbo. Pues bien, al primer subgrupo de sujetos del grupo crítico se les presenta por el canal desatendido, y a la par que esta oración ambigua atendida, la oración (ii): "the spy extinguished the torch in the window" ("el espía apagó la antorcha en la ventana"). Al segundo subgrupo de sujetos se les presenta, también por el canal desatendido y a la par que la oración ambigua atendida, esta otra: "the spy showed the torch from the window" ("el espía enseñó la antorcha desde la ventana"). La tarea experimental encomendada a los sujetos era la de parafrasear el significado de la oración atendida (i). Los teóricos que diseñaron el experimento estaban interesados en determinar si el significado de las oraciones desatendidas tenía una influencia decisiva en la forma que los sujetos elegían a la hora de hacer la paráfrasis del significado de la oración atendida. Si éste fuera el caso, dicha influencia debería manifestarse en la paráfrasis de (i) que los sujetos eligieran.

Para comprobar esta hipótesis, un tercer grupo de control atiende también a la oración (i), pero en este caso el mensaje acústico que se emite por el canal desatendido tiene un contenido semántico tal que no

puede servir en ningún caso de información relevante para eliminar la ambigüedad significativa característica de (i). La hipótesis del experimento dice, más exactamente, que si el significado de las oraciones desatendidas que servirían de contextos en condiciones normales para eliminar la ambigüedad significativa de la oración (i) no consigue ser representado de algún modo por los sujetos del grupo crítico, entonces los patrones de respuestas de los sujetos que se encuentran en esta situación habrán de ser los mismos que los de los sujetos del grupo de control. En caso de que dicho significado sí consiguiera ser representado, los patrones de respuestas deberían ser significativamente diferentes y manifestarse en la paráfrasis elegida por los sujetos del grupo crítico. Los resultados experimentales confirman esta última predicción. Aunque el fenómeno de interferencia sea tan eficaz que los sujetos del grupo crítico no consigan, al final del ensayo, informar acerca de cuál es el contenido semántico del canal desatendido, sin embargo la influencia de ese contenido es incontestable a juzgar por la paráfrasis que eligen de la oración atendida.

Con este experimento como transfondo se facilita la tarea de observar el funcionamiento del modelo explicativo que se esbozó en los primeros párrafos de este capítulo, así como la de entender de qué forma los supuestos de las situaciones experimentales se dan la mano en él. Por lo que respecta a estas últimas, primeramente hay un cierto abismo entre lo que el teórico sabe acerca de la situación experimental global y lo que saben los sujetos de experimentación. En concreto, quien diseña el experimento tiene un conocimiento de las características de los estímulos lingüísticos que se presentan a los sujetos de experimentación del que éstos no disponen. El psicólogo cognitivo, por ejemplo, conoce los fenómenos de interferencia que intervienen en el diseño del experimento. El sabe también cuándo el contenido semántico del canal desatendido tiene, en condiciones normales, una relación significativa con el contenido semántico

del canal atendido y cuándo no. Este es, precisamente, el criterio hetero-fenomenológico que utiliza el psicólogo para distinguir el grupo crítico del grupo de control. También sabe cómo el significado de las oraciones desatendidas, que funcionan de contextos para eliminar la ambigüedad semántica inherente a la oración atendida, puede influir en las paráfrasis elegidas por el grupo experimental; comparte el mundo lingüístico de los sujetos experimentales, de los sujetos psicológicamente semejantes a él. Los sujetos de experimentación, por el contrario, no tienen acceso consciente a esta clase de información. Ahora, el teórico cognitivo trata de "sorprender", por así decir, el funcionamiento y propiedades de la mente computacional, metodológica, con esta clase de experimentos. El psicólogo (el psicolingüista) formula los hechos condicionales, los iffy facts, a que dan lugar las correlaciones funcionales halladas, de la siguiente manera: si en un cierto tiempo *t* se dieran ciertas condiciones (prestar atención al mensaje que se oye por el lado derecho, fenómenos de interferencia en el oído izquierdo), y se diesen ciertos estímulos (emisión de una oración de significado ambiguo en el oído derecho, emisión por el oído izquierdo de oraciones y mensajes acústicos que o bien contribuyen a eliminar la ambigüedad de aquella oración, o bien no contribuyen), se daría cierta respuesta (esta o aquella paráfrasis o ninguna en especial), *ceteris paribus*. A partir de este momento se pasa al ejercicio efectivo del método postulacional que utiliza modelos lingüísticos. No se puede entender las correlaciones observadas si no se postula que el significado de las oraciones desatendidas alcanza un grado de representación semántica suficiente en el sistema lingüístico del sujeto, que hace que éste opte por una u otra paráfrasis.

Hasta aquí, en síntesis, llega la formulación de las hipótesis y el diseño de los experimentos pertinentes para intentar confirmarlas, así como los resultados que se desprenden de esta contrastación. Desde este momento

en adelante el interés se dirige hacia los compromisos ontológicos con los que, en principio, estamos en deuda cuando nos enfrentamos con estos sorprendentes resultados. Lo que voy a decir a partir de ahora es suficientemente bien conocido². Cuando comenta los resultados de este experimento, Fodor afirma que la conclusión a la que guían los datos es que, de alguna manera, "parte de la información sobre el contenido semántico de la oración no atendida se computa <y representa> aun cuando el sujeto no pueda disponer de ella de forma consciente" (Fodor 1975, 162). Sin embargo, la paráfrasis elegida ha de ser una función de esta computación y representación. Pero no puede haber representación del significado de las oraciones, argumenta Fodor, si no hay un "lenguaje" (interno) en el que llevar a cabo esa representación. Por lo demás, este lenguaje, que no puede confundirse en ningún momento con el lenguaje natural en que el sujeto expresa conscientemente el contenido de sus paráfrasis, habrá de ser tan rico y articulado como los propios lenguajes naturales, dado que cada oración del lenguaje natural puede ser adecuadamente representada en él en diversos niveles. Este "lenguaje" o "código" es, por otra parte, inconsciente y, sin embargo, es en este nivel de representación inconsciente en el que se comparan--se computan--los significados de la oración atendida y de la oración desatendida, y en donde se "ejerce la influencia" del significado no atendido sobre el significado atendido. Tanto la oración atendida como la no atendida deben alcanzar un grado de representación semántica en este lenguaje "común" a ambas, de tal forma que la paráfrasis consciente del sujeto es simplemente una función del grado de comparación alcanzado en este nivel de representación "psicolingüística" inconsciente. Por último, no hay manera alguna de

² La fuente clásica para el análisis de las consecuencias de éste y otros experimentos es Jerry A. Fodor. 1975. The Language of Thought. (Cambridge Mass.: Harvard University Press).

adentrarse en el formato y características de este "lenguaje del pensamiento" o "código interno"--así se llama genéricamente a este sistema representacional--si no es entendiéndolo sobre el modelo análogo de los formatos y características lingüísticas de los enunciados de los lenguajes naturales.

Como mencioné en los primeros párrafos de este capítulo cuando comentaba la importantísima obra de Sellars, los enunciados de los lenguajes naturales son la expresión de los "movimientos" relativos de los pensamientos, de aquellos pensamientos sobre cuya naturaleza intrínseca no conocemos conscientemente nada, excepto los heraldos que nos mandan y cuyas expresiones conscientes son los enunciados de los lenguajes naturales que emitimos. En este experimento, por ejemplo, hay "movimientos" relativos entre el pensamiento (i) "the spy put out the torch as our signal to attack" y el pensamiento (ii) "the spy extinguished the torch in the window". Tal y como decía al exponer ideas de Sellars tanto acerca del método de la psicología, como acerca de las propiedades de las entidades postuladas por esta ciencia, los pensamientos son objetos que entran en relaciones análogas--en concreto, aquí, en relaciones semánticas análogas--a la forma en que los enunciados de los lenguajes naturales están relacionados semánticamente entre sí y con los contextos de uso. En tanto que los enunciados de los lenguajes naturales sirven de modelo para la construcción teórica, su carta de presentación, cuando se utilizan para esclarecer las propiedades del lenguaje del pensamiento, es, simplemente, la matización de que el lenguaje del pensamiento es un lenguaje "que no se habla", matización que, por lo demás, puede disuadir a cualquiera de que algo así sea realmente un lenguaje.

Junto a éstas, nuevas recopilaciones de las exposiciones previas se imponen otra vez. El fundamento de la inteligibilidad que da pie al diseño del experimento mismo, reside por entero en el significado ya comprendido

de los enunciados de los lenguajes naturales, significado que comparte el teórico que diseña el experimento con el sujeto de experimentación. Para ambos la oración (i) tiene un significado ambiguo. Para ambos las oraciones (ii) y (iii) servirían de contextos para eliminar la ambigüedad. Sin embargo, el fundamento ontológico necesario para explicar las correlaciones funcionales observadas habrá de ser un nivel de representación inconsciente del significado y de la sintaxis de las oraciones, nivel que requiere la postulación de la existencia de un "lenguaje del pensamiento" que no se habla y en el que se llevan a cabo las computaciones y comparaciones relevantes. Es mediante razonamientos de este tipo, que se apoyan en experimentos concretos de psicología cognitiva, como vamos construyendo paso a paso las propiedades de la "mente computacional" o "mente metodológica", sus partes o módulos más relevantes, así como las interacciones, de naturaleza muy compleja y lábil, entre sus distintos módulos. Sin esta clase de experimentos sería imposible dar con estas propiedades. Ellas no son accesibles a la introspección del sujeto, sino que deben ser postuladas si es que se quiere explicar las correlaciones halladas en el plano observacional.

Para finalizar la exposición y el análisis de las consecuencias de este primer experimento es importante señalar que la utilización que el sujeto hace de este sistema representacional complejo depende en gran medida de cuál sea la tarea que se le plantee en cada caso. De todas maneras, el sujeto, piensa Fodor, tiene que tener acceso a los recursos que pone a su disposición este lenguaje del pensamiento, y los explota de una forma muy flexible en función de las tareas que se le asignan en cada caso y de las situaciones ambientales concretas (Fodor 1975, 194).

(b) Las post-imágenes que menguan.

Con ser sumamente interesante esta clase de consideraciones psicolingüísticas para argumentar en favor de la postulación del lenguaje del pensamiento, existen nuevas evidencias procedentes del ámbito de la psicología que refuerzan y amplían su necesidad. Uno de los experimentos más interesantes y claros que se pueden recordar a este respecto, es el de las postimágenes que menguan (Shrinking After-Images)³. Si hubiera una mancha en la pared, un sujeto que se fuera acercando progresivamente a ella tendría una experiencia consciente subjetiva que describiría diciendo algo así como que le parece que la mancha se hace cada vez más grande. La explicación psicofísica de por qué el sujeto describe así su experiencia es que en la retina se forma una imagen cada vez más grande a medida que se va acercando a la pared. (El tamaño de la imagen en la retina del sujeto aumenta o disminuye en una relación proporcional a la distancia entre él y la pared.) Por supuesto, éste no sería el caso si la mancha de la pared se fuera haciendo cada vez más y más pequeña a medida que el sujeto se acercara a ella. En esta circunstancia, si se consiguiera que la imagen de la mancha que se forma en la retina fuera de un tamaño constante, el sujeto diría que ve la mancha más pequeña.

Supongamos ahora que ese sujeto padece un fogonazo producido por el "flash" de una lámpara y experimenta la percepción consciente de la consabida postimagen. Si mira a continuación a una pared o a una pantalla verá en ella dicha postimagen. En estas circunstancias, se instruye al sujeto para que se acerque progresivamente a la pared. Cuando lo hace así, tendrá una experiencia consciente cuya contraparte lingüística sería: "veo una mancha en la pared que va disminuyendo de tamaño". Es decir, que la

³ Véase Irvin Rock, 1985, (La Percepción, Barcelona: Labor): 25; Richard L. Gregory, 1965, (Ojo y Cerebro, Psicología de la Visión, Madrid: Guadarrama): 153.

postimagen le parece al sujeto más grande cuando la mira como si estuviera lejos y más pequeña cuando la mira como si estuviera cerca. Sin embargo, se puede constatar empíricamente que el tamaño y la forma de la postimagen que el "flash" ha producido en la retina permanecen en todo momento constantes mientras él se acerca a la pared. Se le puede incluso informar de este hecho. Sin embargo, sigue describiendo su experiencia de la misma manera que antes; como una mancha que disminuye de tamaño a medida que se acerca a la pared. En esto, precisamente, consiste la ilusión de las postimágenes, en que parecen variar de tamaño a pesar de que la imagen en la retina es constante. Esta inversión de la relación normal entre el tamaño y la distancia se conoce como "ley de Emmert".

¿Como explicar estos hechos? ¿Qué es lo que debemos suponer que sucede en la mente (metodológica) del sujeto, en este caso en su sistema perceptual? Siempre que la imagen de una mancha sobre la retina tenga el mismo tamaño, sea cual sea su causa, el sujeto describirá lingüísticamente su experiencia consciente como una mancha que va disminuyendo de tamaño a medida que se va acercando a la pared, y ello tanto si hay realmente una mancha en la pared que va disminuyendo de tamaño, como si no la hay y el sujeto, por tanto, padece una ilusión óptica. Descriptivamente, fenomenológicamente, ambos casos son idénticos para el sujeto: la descripción de su vivencia perceptiva permanece idéntica. Heterofenomenológicamente no lo son, sin embargo, dado que en el segundo caso no hay mancha alguna en la pared. Una forma sencilla e "irresistible" (Cummins 1984, 83), por no decir obligada, de explicar esta clase de datos psicológicos, es la postulación de un proceso inferencial que, partiendo de las premisas de que la imagen en la retina tiene el mismo tamaño y de que la distancia entre la imagen y la pared es progresivamente menor, concluye que la mancha disminuye de tamaño.

Cummins extrae numerosas consecuencias de gran valor a partir de esta clase de interpretaciones (1984, 84-85). Me interesa destacar sobre todo dos. La primera es que esta clase de procesos inferenciales ha de transcurrir de una manera completamente inconsciente para el sujeto. Aunque el sujeto del experimento sepa positivamente que no hay ninguna mancha en la pared y que, por tanto, se trata de una postimagen, ese hecho no hace cambiar en ningún rasgo esencial la descripción fenomenológica de su experiencia. Si este proceso inferencial fuera consciente para el sujeto--aquí consciente sólo quiere decir que el sujeto tiene acceso inmediato a él por reflexión y puede "manipularlo"--, entonces esta clase de información actuaría como muro de contención que impediría que el sujeto tenga la clase de experiencia subjetiva que tiene. Sin embargo, nada de esto sucede. El proceso inferencial es completamente impermeable a las creencias explícitas del sujeto. El lema de esta consecuencia es que algo que no es la conciencia del sujeto tiene que "llevar a cabo" este proceso inferencial. Si no el sujeto a través de sus intenciones y creencias explícitas (conscientes), habrá de ser algo así como las creencias implícitas de su sistema visual. Este "lleva a cabo" esta función inferencial de una manera independiente de las creencias explícitas del sujeto. La información que el sistema visual tiene no puede usar la información que el sujeto posee.

La segunda consecuencia es que la experiencia subjetiva cuya contraparte lingüística es "veo una imagen en la pared que disminuye de tamaño", ha de considerarse como "una función de la conclusión que el sistema visual--<seguramente en unión con otros elementos del sistema cognitivo del sujeto y de su estructuración>--produce" (Cummins 1984, 85). Podríamos expresar esta misma idea así: hay una función $I()$ que correlaciona los valores de las conclusiones que "extrae" el sistema visual con las experiencias subjetivas conscientes. $I()$ es una función de interpretación. Según esto, sería el caso que $I(c)=e$; es decir, la experiencia

subjetiva consciente (e) cuya contraparte lingüística es "veo una mancha de en la pared que disminuye progresivamente de tamaño", es una función de la "conclusión neuronal" que el sistema visual extrae en virtud del aparato inferencial con el que viene equipado innatamente.

(c) **"Aprendizaje" temporal visomotriz.**

El tercer ejemplo sobre el que quiero llamar la atención es un experimento de Finke (1979 y 1980). A un sujeto de experimentación se le ponen unas gafas con cristales prismáticos. El efecto inmediato que produce este tipo de gafas es el de desplazar el campo visual hacia un lado, de tal modo que un sujeto al que se le han colocado este tipo de gafas manifestará claros errores de coordinación visomotriz. Mientras que antes era perfectamente capaz de señalar y coger un objeto que estaba delante de él, ahora, con las gafas puestas, apuntará equivocadamente hacia un lado, equívoco del que él puede llegar a tener clara conciencia. Simplemente no atina a coger los objetos que ve. El experimento consiste en dejar que el sujeto compruebe y señale de nuevo, sin las gafas puestas, donde está el objeto que ha señalado antes de una forma desplazada cuando las tenía puestas. "Retroalimentación visual" se llama técnicamente a esta manera de proceder. Tras varias pruebas de este estilo se observa que el sujeto es capaz de indicar cada vez más y más exactamente dónde está el objeto. Tras someterse a este entrenamiento durante un tiempo se le quita definitivamente las gafas para que pueda recuperar su "visión normal". El efecto observado que se produce cuando se hace esto es que ahora el sujeto señala los objetos con un desplazamiento en sentido contrario al anterior. Si antes, con las gafas puestas, señalaba, por ejemplo, hacia la izquierda del objeto debido al efecto de desplazamiento que producen las gafas sobre el campo visual, ahora, tras el entrenamiento, señalará hacia su

derecha. "Este fenómeno indica un proceso de adaptación visomotriz a una distorsión producida en el campo visual" (Vega 1984, 241). No obstante, es importante señalar el dato de que si no hay retroalimentación después de cada uno de los ensayos no se da el efecto observado.

¿Cómo se podría interpretar este curioso efecto investigado por Finke? Desde mi punto de vista habría que decir respecto de él algo muy semejante a lo que se dijo en los casos anteriores. Una explicación sencilla y coherente de los efectos comportamentales observados en el sujeto, es que su sistema visomotriz ha "aprendido" y "memorizado" una determinada pauta de coordinación que se podría resumir toscamente con la expresión: "lo que el ojo dice al brazo que señala". Dado que tras el entrenamiento y durante un periodo de tiempo breve, el sujeto señala sistemáticamente en sentido contrario al que señalaba antes (cuando tenía las gafas puestas), parece que el sistema visual—ciertos mecanismos de procesamiento visual—tiene que tener alguna forma de "decir" al brazo que él ha "aprendido"—supuestamente gracias a su coordinación con los movimientos del propio brazo—que el objeto está desplazado siempre hacia un lado, que lo ha de buscar allí. Pero no hay forma de "decir" nada ni de "aprender" nada si es que no hay un sistema representacional en el que poder expresar la información "aprendida", en este caso, la información: el objeto está desplazado hacia un lado, señala hacia allí. El movimiento intencional del brazo por parte del sujeto, sería solamente la función de lo que el sistema visual ha "aprendido" y "dicho" al sistema motriz de una forma completamente inconsciente para el sujeto.

& 8. Compromisos ontológicos.

Con estos tres ejemplos a la vista empieza a emerger una cierta teoría acerca del ser psíquico que es, desde hace dos o tres décadas, la

reina del lugar y que se conoce, desde que su principal mentor, Jerry A. Fodor, la bautizara así, como "Teoría Representacional de la Mente"; (a partir de ahora me referiré a ella de una forma abreviada como TRM). La TRM, tal y como muestran los pasos que han conducido hasta ella, no es una teoría de primer orden acerca de cierto grupo de fenómenos--los fenómenos propiamente psicológicos--, sino una teoría de segundo orden acerca de las bases teóricas de la psicología; una teoría, por consiguiente, que habla también acerca de cuáles son los compromisos ontológicos de la psicología cognitiva. Los detalles de la TRM son de sobra conocidos, y sólo los enunciaré a vuela pluma⁴. La TRM dice que tiene que haber un sistema representacional lo suficientemente rico, potente, variado y estructurado en sus capacidades representativas como para poder expresar en él, por ejemplo, el contenido significativo de las oraciones, las inferencias, los procesos involucrados en el aprendizaje y, en general, un gran número de operaciones psicológicas, tal y como requiere la clase de explicaciones a la que parecen conducirnos los fenómenos psicológicos observados. La TRM dice que tiene que haber un lenguaje del pensamiento; un lenguaje en el que se realicen las variadas operaciones--de tipo computacional--que exigen las explicaciones psicológicas. En general, la tesis última de la TRM sostiene que la exhibición de comportamiento intencional es la forma acabada y observable del entrecruzamiento de las variadísimas operaciones que se llevan a cabo en este sistema representacional. En general, cualquier actividad y capacidad cognitiva desplegada por el hombre--o por cualquier otra clase de ser que exhiba capacidades cognitivas semejantes--, es el signo visible de complejos procesos computacionales, inconscientes, "internos"

⁴ Las siguientes obras son algunas de las fuentes clásicas fundamentales que pueden consultarse en relación con este tema: Gilbert Harman. 1973. Thought. Princeton: Princeton University Press. Jerry A. Fodor. 1975. The Language of Thought. Cambridge, Mass.: Harvard University Press; Jerry A. Fodor. 1981. Representations. Philosophical Essays on the Foundations of Cognitive Science. Cambridge, Mass.: The MIT Press; Hartry H. Field. 1978. "Mental Representations." Erkenntnis 13: 9-61.

(ocultos a la percepción tanto externa como interna) y, probablemente, innatos, que subyacen y posibilitan la manifestación de aquellas capacidades y actividades. Pero no solamente esto. La TRM también afirma que cada vez que hay un estado mental subjetivo con un determinado contenido proposicional, esto es, cada vez que un sujeto está en una determinada actitud proposicional, cada vez que un sujeto percibe un objeto, lo imagina, cada vez que cree o recuerda algo, dicho sujeto está también en una determinada relación con una *fórmula* de este sistema representacional en la que el contenido de cada una de sus actitudes proposicionales es el que es porque el contenido de la fórmula es el que es, de tal forma que la naturaleza de una actitud proposicional consistirá, de una manera resumida, en lo siguiente:

Para cualquier organismo *O* y para cualquier actitud *A* hacia la proposición *P*, hay una relación *R* ("computacional/funcional) y una representación mental *RM* tal que
RM significa que *P* y
O tiene *A* sii *O* mantiene una relación *R* con *RM*. (Fodor 1987, 17; véase también Fodor 1975, 75).

Es decir, que exhibir una capacidad cognitiva consiste en ejecutar una de estas complejas computaciones. No hay ningún plus en la manifestación de la capacidad cognitiva que no esté contenido en el proceso computacional mismo considerado como un todo. Esta tesis, de acuerdo con Fodor, expresa cuál es el núcleo central de la ontología con la que está comprometido el programa de investigación de la psicología cognitiva y, en general, el de las ciencias cognitivas (Fodor 1975, 199; Fodor 1987, 17-18; Lycan 1981, 16-19). Precisamente, el programa de investigación razonable que ha encontrado la psicología cognitiva contemporánea es caracterizar y articular en sus diversas partes y niveles este sistema representacional.

Como se ha podido apreciar, llegar hasta la formulación de los compromisos ontológicos de la psicología cognitiva, al menos tal y como los ven Fodor y muchísimos otros autores, supone andar un largo camino lleno de dificultades y asperezas argumentativas. El lector atento y paciente, sin embargo, estará pendiente todavía de la satisfacción de un último deseo: el de saber qué tienen que ver estos compromisos ontológicos de la ciencia psicológica cognitiva con el ser psíquico y con las propiedades que le atribuimos desde el punto de vista del sentido común. En el capítulo anterior se puso de relieve que el ser psíquico es una especie de "anormalidad" ontológica, en el sentido de que es algo que puede entrar en relaciones causales--como lo hacen las bolas de billar--, pero que, sin embargo, tiene y preserva al mismo tiempo relaciones semánticas características--como lo hacen las oraciones de los lenguajes naturales y las proposiciones. A esta propiedad, como se recordará, se la calificó como la más profunda y enigmática de la mente. Ahora se requiere una explicación que ponga en armonía las dos partes que la componen--la parte causal con la parte intencional--, armonización que constituye el reto permanente que la reflexión sobre el fundamento inteligible de la psicología pone a toda investigación acerca de su fundamento ontológico último. También en el capítulo previo se expresó que un pensamiento listo para reflexionar acerca de "la" psicología popular encontraría tarde o temprano la conveniencia de distinguir entre estados mentales representacionales y procesos mentales. Los procesos mentales eran cadenas causales cuyos eslabones ineliminables serían los diferentes estados mentales intencionales. La vida psíquica completa estaría formada esencialmente por los componentes representacionales de estas cadenas causales y por sus modos de unión causales característicos. El comportamiento intencional es sólo la "manifestación" del extremo último de cada una de estas cadenas causales que funcionarían, verdaderamente, como sus causas eficientes. Ahora la pregunta es: ¿cómo integrar este

cuadro con los compromisos ontológicos de la psicología científica? Es decir, ¿qué idea feliz, científicamente tratable, puede ofrecer la conexión deseada entre "la" psicología popular y los compromisos ontológicos de la psicología objetiva? ¿Qué relación existe entre las actitudes proposicionales, tal y como las contruye el sentido común, y las fórmulas del lenguaje del pensamiento? La respuesta a esta pregunta es--¡una vez más!--tan controvertida como cualquiera de las que se dieron anteriormente a otras cuestiones. Una muestra. Quien sea un materialista eliminativista pensará que no hay razón alguna para comprometerse con la clase de entidades que nos ofrece la reflexión sobre el ser psíquico, al menos tal y como se nos da éste en el mundo de la vida. Desde este punto de vista, como no hay nada que tenga a la vez propiedades intencionales y causales porque toda la ciencia que se necesita para explicar causalmente el comportamiento inteligente es la neurofisiología, no hay ninguna necesidad de ponerse a buscar de qué manera un ser dotado de semejantes propiedades se relacionaría con los compromisos ontológicos de la psicología científica. Sin embargo, sí que hay algo que prácticamente todo el mundo está dispuesto a afirmar: si la investigación psicológica de esta segunda mitad de siglo puede contribuir de una forma relevante a solucionar algunos problemas filosóficos clásicos, entonces su granito de arena es, precisamente, esta idea feliz (Cummins 1984, 52; Fodor 1987, 18). Todo lo demás es un viejo conocido.

El "truco" que ofrece el puente de unión consiste en hacer uso abundante de la "metáfora del ordenador". Un ordenador es un sistema físico bastante complejo cuya propiedad esencial es la de tener mecanismos que pueden ser explotados para poner en correlación las propiedades semánticas de ciertos "símbolos" con sus propiedades causales a través de sus propiedades sintácticas (Fodor 1987, 18-24). Damos a un ordenador la orden de imprimir un texto en la pantalla utilizando el lenguaje de

programación BASIC. Tecleamos, por consiguiente, los caracteres PRINT, dejamos un espacio en blanco y a continuación, entre comillas, escribimos el texto que queremos imprimir. Si la orden respeta la sintaxis del lenguaje de programación y todo lo demás va bien, entonces conseguiremos imprimir ese texto en la pantalla. Si la sintaxis no es correcta, nuestro propósito se verá truncado. El propósito que queremos alcanzar está esencialmente mediado, por consiguiente, por la sintaxis del lenguaje de programación. El ordenador "obedecerá" de tal forma que aquello que hemos puesto en él obrará para que, al final, aparezcan ciertos caracteres en la pantalla. Pero, ¿qué es lo que hemos puesto en él? Si la sintaxis es correcta, entonces el mensaje PRINT, gracias a un lenguaje de compilación de BASIC, se "traducirá" en una determinada secuencia de estados físicos de la máquina que no son otra cosa que una determinada "fórmula" del lenguaje máquina (para más detalles, véase, por ejemplo, Fodor 1975, 66-67). A cada mensaje del lenguaje de programación que esté sintácticamente bien formado le corresponde un estado físico complejo de la máquina. Desde este punto de vista, aquello que hemos puesto en el ordenador cuando le damos una orden con un significado determinado es, con toda seguridad, un estado físico más o menos complejo. Ahora comienza la computación de este mensaje. Aquel estado físico causa ciertos otros estados físicos en una cadena cuyo eslabón último es la obtención del efecto deseado: que en la pantalla aparezcan impresos ciertos caracteres. A la luz de esta clase de consideraciones, el estado físico que hemos puesto en la máquina parece que podemos y debemos interpretarlo como la orden sintácticamente bien formada que significa "imprime esta cadena de caracteres en la pantalla". Una fórmula en el lenguaje máquina es, ciertamente, un estado físico, pero también es un "símbolo" que tiene propiedades sintácticas y semánticas. Paralelamente, los estados físicos que este estado físico ha causado debemos interpretarlos no como una mera cadena de estados físicos causalmente

activos y así interrelacionados, sino como pasos a lo largo de una computación. El programa está corriendo.

Y ahora la comparación y aplicación de este "truco" al caso de la psicología y a su objeto de estudio.

Cuando pensamos en un organismo como un computador, intentamos asignar fórmulas en el vocabulario de la teoría psicológica a estados físicos del organismo (por ejemplo, a estados de su sistema nervioso). Idealmente, la asignación se debería llevar a cabo de tal manera que (al menos algunas) de las secuencias de los estados causalmente implicados en la producción del comportamiento se puedan interpretar como computaciones que tienen descripciones apropiadas del comportamiento como su "última línea"... La idea es que, tanto en el caso de los organismos como en el caso de los computadores, si conseguimos la forma correcta de asignar fórmulas a los estados, será factible interpretar los sucesos que *causan* las salidas (*outputs*) como una *derivación* computacional de las salidas. Brevemente, los sucesos orgánicos que aceptamos como implicados en la etiología del comportamiento resultarán tener dos descripciones teóricamente relevantes si las cosas salen bien: una descripción física en virtud de la cual caen bajo leyes causales y una descripción psicológica en virtud de la cual constituyen pasos en la computación desde los estímulos hasta las respuestas. (Fodor 1975, 73-74).

Una y otra vez aguarda la labor de unificar poco a poco los resultados que se van obteniendo para conseguir un cuadro teórico unitario. Un nuevo elemento valiosísimo se incorpora ahora en este estadio. En la discusión pormenorizada de las críticas al fisicalismo positivista hubo ocasión de analizar cuál era la doble descripción que admiten los sucesos involucrados en la causación de la conducta. Estos eran sucesos que satisfacen descripciones físicas, gracias a las cuales caen bajo leyes físicas, pero que también satisfacen descripciones psicológicas. La presencia de la TRM ofrece ahora algunas ideas sustanciales acerca de en qué consistirían más exactamente las descripciones psicológicas de estos sucesos físicos causalmente activos. Hay que tener en cuenta, en cualquier caso, que lo

que diga la TRM acerca de este punto debe ser compatible con lo que se dijo en el párrafo 5; compatible, sobre todo con la tesis de que los estados psicológicos supervienen en los estados físicos. La analogía con el funcionamiento y las características de los ordenadores posibilita la comprensión de esta doble descripción, sin que se viole la superveniencia. En efecto, una computación es un suceso peculiarísimo que tiene lugar en ciertos sistemas físicos complejos y que satisface tanto descripciones físicas, como descripciones semánticas y sintácticas. Pero, como puede fácilmente comprenderse, cualquier computación superviene necesariamente en los estados físicos del sistema, de tal manera que una mera descripción física de éste sería insuficiente para encontrar posteriormente nada semejante a un "símbolo" computacionalmente relevante y semánticamente interpretable. Gracias a la superveniencia de un estado computacional en un estado físico, los "símbolos" tienen capacidad causal, porque la tienen los estados físicos sobre los que supervienen. De una manera paralela, un estado mental es algo que satisface también una descripción física y una descripción semántica. En virtud de su descripción semántica, los estados mentales son intencionales y tienen condiciones de satisfacción--son estados propiamente representacionales--, pero en virtud de su descripción física los estados mentales entran en relaciones causales. Los estados mentales son también, en este sentido, "símbolos"⁵. Gracias a estas relaciones causales, los estados mentales intencionales pueden entrar en las cadenas causales que dan origen a las conductas intencionales. Por este motivo, hay fuertes razones teóricas y empíricas para pensar en la vida psíquica como vida computacional. Las fórmulas del lenguaje del pensamiento, a las que Fodor llama también en

⁵ Es claro que la tradición reciente en filosofía de la mente no utiliza el término "símbolo" en el sentido inaugurado por C.S. Peirce, es decir como algo que designa o representa arbitrariamente. Los símbolos del Lenguaje del Pensamiento no podrían representar arbitrariamente, aunque, por razones que se analizarán más adelante, parece que deberían poder representar falsamente. Los organismos, sobre todo los humanos, se equivocan y cometen errores.

muchas ocasiones "representaciones mentales" o "símbolos" no son otra cosa que estados físicos sumamente complejos del organismo y que deben ser interpretados semánticamente y que tienen, por lo tanto, capacidad representacional. Crudamente expresado, una representación mental es un estado físico que significa, un estado físico que es objeto de evaluación semántica. De una forma complementaria, un proceso mental, una cadena causal de estados mentales, es simplemente una secuencia causal de representaciones mentales. Otra forma de expresar esto mismo sería diciendo que la TRM está comprometida con el *realismo estándar*. La idea básica del realismo estándar es que las actitudes proposicionales forman redes causales e inferenciales que son isomórficas (Fodor 1985, 86). Las actitudes proposicionales tienen esencialmente capacidad causal y en su interacción forman complejas redes causales, de tal manera que cada actitud proposicional concreta ocuparía un nodo en esta múltiple red causal. Es así, gracias a las localizaciones de las actitudes proposicionales en los nodos de estas redes causales, como se contaría con un criterio imprescindible para individualizarlas.

Para cada estado mental hay un papel causal correspondiente, y para cada papel causal hay un nodo correspondiente (Fodor 1985, 85).

Si es convincente la cadena de razonamientos que ha conducido hasta aquí, entonces no contamos con ningún elemento que nos indique cómo se podría individualizar una actitud proposicional si se excluyen y marginan sus relaciones causales con otras actitudes proposicionales. Esta es la tesis clásica del funcionalismo, que enseña que los estados mentales son estados funcionales, es decir, estados que se individualizan por referencia a sus relaciones y papeles causales con estímulos, con respuestas y, sobre todo, con otros estados mentales. Los estados mentales, pues, no tienen "esencias físicas", sino "esencias funcionales" (Lycan 1981, 15). En el caso más simple,

toda actitud proposicional tiene un halo actual o potencial de relaciones causales con otras actitudes proposicionales. Este halo contribuye a su individuación-tipo. Ser una percepción, ser una imaginación o ser la comprensión de una oración, son tipos de actitudes proposicionales que se individúan por referencia a sus papeles causales-funcionales, esto es, por referencia a los lugares que pueden ocupar en esta red causal, y, por lo tanto, por referencia a las potencialidades de interacción causal con otras actitudes proposicionales. En otras palabras, los papeles causales de una percepción y de una imaginación no pueden ser iguales, porque sus halos potenciales de interacción causal con otros tipos de estados mentales son diferentes, y esto es así tanto si tienen como objetos las mismas proposiciones, como si no (Fodor 1985, 85; Fodor 1987, 17). Percibir una manzana roja e imaginarse una manzana roja son dos actitudes proposicionales que no pueden individuarse simplemente apelando al hecho de tener una manzana roja como su objeto proposicional, dado que en este respecto concreto son idénticas. Según el funcionalismo, si son dos actitudes proposicionales de tipos diferentes es porque tienen, actual o potencialmente, distintos patrones de relaciones causales con otras actitudes proposicionales. Las constelaciones potenciales de interacción causal de las percepciones y las constelaciones potenciales de interacción casual de las imaginaciones son diferentes. Si no hubiera ninguna diferencia en este respecto, percibir una manzana roja e imaginarse una manzana roja serían dos actitudes proposicionales esencialmente idénticas.

Sin embargo, Fodor considera que individuar los tipos de actitudes proposicionales por referencia a sus papeles causales-funcionales no es suficiente para precisar por qué, por ejemplo, una percepción determinada tiene el contenido proposicional que de hecho tiene; por qué esta percepción lo es precisamente de esta manzana roja (Fodor 1985, 85). No cabe duda de que el contenido proposicional de una actitud debe contribuir

decisivamente a su individuación plena. Aunque las percepciones tengan, característicamente, halos potenciales de relaciones causales que son diferentes de los halos propios de las imaginaciones, con todo, no se podrá negar que la percepción de esta manzana roja y la percepción de esta grapadora azul tienen que tener también distintos halos potenciales de interacción causal. Como se suele expresar intuitivamente en algunas ocasiones, aunque una y otra percepciones estén en la misma "caja" (la caja de las percepciones), sus objetos proposicionales hacen que no entren en exactamente las mismas relaciones causales con otras percepciones o con otras actitudes proposicionales. La percepción de una manzana roja puede entrar en relación causal con mi deseo de comerme una, y hacer finalmente que coja una manzana roja y me la lleve a la boca, *ceteris paribus*. No así puede hacerlo mi percepción de una grapadora azul. Por consiguiente, dado que las actitudes proposicionales tienen proposiciones como sus objetos, podemos utilizar también estas proposiciones y las relaciones semánticas de inferencia que se dan mutuamente entre ellas y que les son inherentes para alcanzar su identificación plena. Es decir, que a partir de este momento hay que contar con dos especies distintas de redes: las redes causales y las redes inferenciales. El isomorfismo entre estas dos especies de redes al que hacía mención más arriba es tal que *"el papel causal de una actitud proposicional refleja (mirrors) el papel semántico de la proposición que es su objeto"* (Fodor 1985, 86). Bien pensado no podría haber sido de otra manera, dado que el papel causal de una proposición y el nodo que ella ocupa en la red causal general dependen no sólo del tipo de actitud proposicional que es, sino también de su objeto proposicional. Es este isomorfismo el que ha motivado que en numerosas ocasiones se haya llegado a pensar que se puede "reducir la noción de contenido proposicional a la noción de papel casual" (Fodor 1985, 85; véase también Cummins 1989, 114). Si se supiera cuál es el papel causal de una actitud proposicional, si se supiera cuáles

son las generalizaciones causales que la subsumen, se sabría correlativamente cuál es su contenido. Sin embargo, esta clase de reduccionismo corre el peligro, si se considera la cuestión con un poco más de detenimiento, de incurrir en círculo vicioso. Todo aquel que defienda que el papel funcional-causal de una actitud proposicional determina su individuación plena, y que considere, por esta misma razón, que una actitud proposicional tiene el contenido que tiene en virtud de su papel funcional/causal, está a un paso de verse obligado a sostener, sobre todo si defiende el realismo estándar, que sólo se puede "*identificar* los papeles funcionales relevantes <de una actitud proposicional> por referencia al estatus representacional del estado <mental> y a su contenido" (Cummins 1989, 116).

La preocupación subyacente que anima toda esta discusión en torno al problema de la individuación de los estados mentales y, en especial, aquella que subyace al interés que muestra el funcionalismo por ofrecer una noción suficiente que permita individuar los estados mentales—noción que se encontraría, según él, en el concepto analizado de "papel causal"—, es ofrecer una teoría de las representaciones mentales que sea compatible con una ontología naturalista. Este es el fondo del saco de los problemas que preocupan hoy en día a la fundamentación de la psicología. La noción de causa es esencial para toda ontología de la naturaleza, para toda ontología de las ciencias naturales. Si la psicología es una ciencia natural, entonces sus conceptos y enunciados se referirán a objetos y estados de cosas que entran en relaciones causales. Dado que la psicología tiene las actitudes proposicionales como sus objetos por excelencia, en tanto que ellas explican causalmente la conducta intencional de los organismos, entonces las actitudes proposicionales son susceptibles de individuación por sus papeles causales. Si toda actitud proposicional tiene una proposición como su objeto; si no hay actitud proposicional cuando no existe una representación mental que tenga el mismo significado que la proposición de la actitud, y si toda

representación mental es un estado físico causal y computacionalmente activo, entonces hay que explicar cómo algo físico puede significar.

El problema que muchos filósofos ven es que, aun cuando sea deseable que la psicología no se salga del armazón de la ontología naturalista, y que, por tanto, sea necesario defender una u otra versión del funcionalismo, todavía hace falta explicar por qué ocupar una determinada posición nodal en una red de relaciones causales hace que una fórmula del lenguaje del pensamiento tenga un significado determinado, por qué el contenido proposicional de una percepción representa lo que representa. Este es un problema con el que habrá que enfrentarse tarde o temprano.

Una nueva recopilación se impone ahora. De acuerdo con la TRM no hay actitudes proposicionales si no hay representaciones mentales. Por otra parte, el objeto proposicional de una actitud es el que es, porque la representación mental significa lo que significa. Harman ha expresado nítidamente la consecuencia esencial de esta tesis: "las características representacionales de los estados mentales derivan de las características representacionales de los enunciados del lenguaje del pensamiento" (1973, 57); y Fodor la ha sentenciado cuando afirma que las representaciones mentales "funcionan tanto de objetos inmediatos de las actitudes proposicionales, como de dominios de los procesos mentales" (1987, 16-17). Según Fodor, precisamente esta clase de afirmaciones son las que diferencian a la TRM respecto del mero realismo intencional. El realismo intencional dice que hay actitudes proposicionales, es decir, estados mentales de creencia, de deseo, estados mentales perceptivos, imaginativos, etcétera, y sostiene que estos estados mentales tienen un contenido intencional. Por otra parte, el realismo intencional también defiende que los estados mentales tienen capacidad causal, y que el comportamiento intencional es el efecto de las múltiples y complejas interacciones entre estos estados mentales. Finalmente, el realismo intencional debe ser necesariamente compatible con

la ontología materialista y fisicalista. Si los estados mentales tienen de verdad capacidad causal, entonces es porque son materiales. Por lo demás, respecto de la ontología fisicalista ya hubo ocasión suficiente de matizar en qué sentido gran parte de la filosofía de la mente contemporánea se quiere alejar de las tesis estrictas del fisicalismo positivista. También hubo oportunidad de discutir las "anomalías" curiosas que surgen cuando se sostiene a la vez una ontología fisicalista en la que los estados psicológicos y los estados físicos son contingentemente idénticos, pero en la que no hay relaciones nomológicas entre unos y otros. Con todo, el fisicalismo se presenta como una ontología respetable también para la psicología. Si la psicología es una ciencia, no podría haber sido de otra manera. Así, pues, lo que la fundamentación de la psicología ha ganado en este siglo para la reflexión filosófica, después de la crisis conductista de los años sesenta, es que se puede defender que la psicología se ocupa de los estados mentales de los organismos, sin por ello tener que traicionar la ontología materialista (Fodor 1987, 135). La postulación de un lenguaje del pensamiento o, lo que es lo mismo, la construcción de teorías psicológicas funcionales/computacionales, pone de manifiesto cómo se podrían armonizar todas estas tesis. Como ya se vio, la TRM propone interpretar las actitudes proposicionales de los sujetos como fórmulas del lenguaje del pensamiento, como muestras de representaciones mentales, de símbolos. El contenido intencional de la actitud proposicional está determinado, a su vez, por el significado de la representación mental, del símbolo. La ontología materialista, por su parte, no se viola con esta clase de teorías representacionales, porque los símbolos mentales son materiales en el sentido de que admiten una descripción física. Sin embargo, los símbolos mentales no se reducen a estados físicos, porque son estados computacionales, funcionales y representacionales del organismo.

Curiosamente, a la luz de esta serie de ideas, los compromisos ontológicos de la psicología cognitiva revelarían que ésta, en algunos aspectos importantísimos, es una psicología profundamente aristotélica.

Precisamente por esto están en lo cierto cuantos opinan que el alma ni se da sin un cuerpo, ni es en sí misma un cuerpo. Cuerpo, desde luego, no es, pero sí algo del cuerpo, y de ahí que se dé en un cuerpo y, más precisamente, en un determinado tipo de cuerpo: no como nuestros predecesores que la endosaban en un cuerpo sin preocuparse de matizar en absoluto en qué cuerpo y de qué cualidad, a pesar de que ninguna observación muestra que cualquier cosa al azar pueda recibir al azar cualquier cosa. (Aristóteles. De Anima. Libro II. Capítulo 2).

Si algo se ha ganado para el conocimiento de la psicología humana en nuestro siglo, al menos por lo que hace a los procesos psicológicos cognitivos más básicos, es que ahora contamos con una buena idea acerca de cómo el alma no es cuerpo pero es algo del cuerpo. Todo lo que se acaba de decir apunta hacia la solución de este problema, acompañante inseparable de la reflexión filosófica durante siglos. En cualquier caso, los aristotélicos deberían mirar con buenos ojos el descubrimiento y aplicación de la "metáfora del ordenador" al caso de la psicología.

La postulación del lenguaje del pensamiento, pues, es la solución relativamente simple y convincente--hasta donde puedan convencer los compromisos ontológicos de las ciencias especiales, que nunca es convicción apodíctica--del problema de la inteligencia encarnada, del antiquísimo problema que preocupaba a Sócrates de cómo la inteligencia (Nous) es causa, del problema kantiano de cómo podría haber causación por medio de meras representaciones. Los estados mentales son representacionales, intencionales, y los procesos mentales son cadenas causales de estados físicos que admiten y, más aún, deben ser interpretados como manipulaciones de esas representaciones, es decir, deben ser interpretados semántica y

computacionalmente. Hay que interpretar, por consiguiente, las cadenas causales desde las entradas (inputs) (de naturaleza completamente variable y que además no se conoce muy bien) hasta las respuestas (outputs) (las cogitaciones o los comportamientos típicamente cognitivos), como manipulaciones de entidades físicas que tienen propiedades semánticas o que son semánticamente interpretables. Resumiendo, el ser psíquico en su fundamento ontológico, la esencia real del ser psíquico que se buscaba desde hace tanto tiempo es, según parece, ésta: una fórmula de este lenguaje interno del pensamiento, una fórmula de este sistema representacional. Así, pues, adoptar la actitud dogmática-natural propia de la filosofía objetivista a propósito del dominio empírico de la subjetividad nos ha conducido provisionalmente hasta aquí.

La actitud filosófica objetivista, cuando se enriquece con los programas de investigación de la psicología actual, dice que debemos tomarnos absolutamente en serio la idea de que los estados intencionales de los organismos tienen una relación peculiarmente íntima con sus estados físicos. Esta tesis, evidentemente, tiene consecuencias de gran alcance. Si hay una relación de identidad contingente entre los estados físicos y los psicológicos, entonces toda teoría que quiera "restar" al análisis de los estados intencionales el análisis físico y computacional será una teoría "ideal" sobre la intencionalidad. Tomarse absolutamente en serio la función del cuerpo a la hora de construir una teoría de la intencionalidad satisfactoria y completa hace que no se pueda eludir la tesis de que estar en un estado intencional determinado, por ejemplo, en el estado que consiste en percibir una manzana, no es sino estar en un determinado estado físico/computacional. Dicho con la mayor crudeza posible. Realmente, no tenemos ni la menor idea de a qué se podría parecer la vida psicológica de un ser sin cuerpo. Si la filosofía de los tiempos pasados ha creído poder construir teorías que niegan lo que acabo de decir, tenemos

buenas razones para pensar que el motivo por el cual se ha procedido de esta manera es la característica "transparencia" del lenguaje del pensamiento. Gracias a ella, parece que da igual tener realmente cuerpo que tener un cuerpo soñado o no tener ninguno. No parece que el cuerpo quite o ponga algo esencial para hacer una teoría de la intencionalidad seria. Y, sin embargo, lo absolutamente esencial ha pasado siempre desapercibido. No es de extrañar. Nunca se hubiera podido percibir de ninguna manera: ni por percepción externa, ni por percepción interna. Sólo se puede postular y dar pruebas constructivistas de su existencia. Me atrevería a decir, a la luz de estas reflexiones incipientes, que nunca como hasta la hora de conocerse teorías semejantes a la TRM se ha puesto al cuerpo en un lugar tan central en la elaboración de una teoría de la intencionalidad.

Esta tesis tiene también, desde luego, consecuencias epistemológicas de gran alcance y que sólo voy a indicar ahora muy de pasada. Debemos pensar que los objetos, tal y como los describimos en la experiencia fenomenológica, nos son dados después de que el sistema nervioso ha pasado por ciertos estados causales que deben interpretarse computacional y semánticamente. No tenemos ni la menor idea de qué podría ser un objeto "en sí" ni qué propiedades podría tener al margen de cómo nos es dado después de atravesar completamente el círculo causal-representacional neurofisiológico. Que parezca que podemos hablar de los objetos al margen de cualquier caracterización de los estados neurofisiológicos y de las computaciones que ellos llevan a cabo se debe especialmente a que todo lo que sucede en este nivel neurofisiológico y computacional es, con toda seguridad, absolutamente inconsciente. Sólo que si no hay computación, no hay intencionalidad. Es decir, no tenemos ni idea de a qué podrían parecerse el mundo y sus objetos si no tuviésemos sistema nervioso, y no cumpliese con las funciones que le son propias. Dudar del cuerpo es la clase de cosas que siempre cabe hacer. La tentación de no contar con el

sistema nervioso, con el cuerpo, se debe a que todo lo que él hace y es relevante y necesario para la percepción de otros cuerpos, pasa igualmente desapercibido.

Por lo demás, la capacidad representativa del lenguaje del pensamiento es tan sumamente perfecta que sólo se puede llegar a tener noticia de ella de una forma muy indirecta, aunque, desde luego, racional: mediante la construcción de las teorías hipotéticas explicativas que son necesarias para dar cuenta de las correlaciones funcionales observadas en cierto dominio empírico: el propio de la psicología. Por un camino absolutamente inesperado en la historia del pensamiento filosófico occidental habríamos topado aquí de nuevo con los *signos formales*, con los signos puros de los que ha hablado el pensamiento escolástico medieval, con aquellos signos que son "tan perfectos que contienen la realidad misma de lo significado....Estos signos, por principio, realizarían tan adecuadamente la naturaleza o forma del signo--estar por otro revelándolo cognoscitivamente--, que no se *interpondrían* verdaderamente entre el cognoscente y lo conocido" (García-Baró, 132). El lenguaje del pensamiento es una transparencia absoluta que, sí, hay que descubrir, pero sin la cual seríamos cognitiva e intencionalmente ciegos. Y con nosotros, cualquier otro organismo del que estemos dispuestos a sostener que tiene vida psíquica.

Y ahora lo que importa saber es cómo es posible explicar la intencionalidad de las representaciones mentales, de los símbolos mentales; cómo consiguen representar "el mundo". Mucho antes que analizar este problema que, según es normal decir, nadie ha conseguido solucionar todavía a la satisfacción de todos, quiero señalar que parece haberse dado un cierto desplazamiento inadvertido desde la afirmación de que ha de haber un sistema representacional en el que poder expresar los significados, inferencias, etc., a la afirmación de que los símbolos de ese sistema representacional *son* los significados, etc. "Los conceptos son fórmulas de

Mentalés" (Fodor 1987, 160). Pero primero dejemos oír hablar la sinceridad intelectual de Fodor para dejar sentada definitivamente la enunciación del problema de la intencionalidad de las representaciones mentales.

Pero el LDP guarda un silencio total--por no decir ensordecedor--sobre un problema fundamental con el que la psicología computacional tendrá que enfrentarse en algún momento: ¿de dónde proceden las propiedades semánticas de los símbolos mentales? Los supuestos metafísicos de LDP son materialistas; se da por supuesto que las representaciones mentales son objetos físicos. Pero en ese caso, ¿cómo pueden tener propiedades semánticas los objetos físicos? ¿Cómo pueden versar los unos sobre los otros? Este problema procede en línea directa del famoso "problema de la intencionalidad" de Brentano; en realidad, es la forma que adopta el problema de Brentano en el contexto de una teoría representacional de la mente. En la actual situación, no tenemos la menor idea de cómo solucionarlo, aunque debemos intentarlo si queremos tener una ciencia cognitiva que sea al mismo tiempo materialista y computacional.... Lo que LDP insinúa--quizá de una forma sorprendente--es que tampoco puede haber una teoría de la *mente* sin una teoría de los símbolos. (1984a, 21).

Tres problemas prioritarios van a ocupar la investigación a partir de este momento. El primero de ellos tiene que ver con la tarea de dilucidar qué clase de estructura tienen las representaciones mentales, si es que tienen alguna. Esta primera serie de investigaciones está especialmente unida a la aclaración de la cuestión de si los objetos de las actitudes proposicionales, las proposiciones, tienen algún tipo de estructura. El segundo de los problemas que se tratará tiene que ver esencialmente con la posibilidad de ofrecer una teoría semántica para estos símbolos mentales que no traicione la ontología materialista, esto es, que no haga uso, como suele exigirse, de nociones intencionales o semánticas. El tercer y último problema es la consideración de lo que sucede en la psicología contemporánea con la conciencia que acompaña a todos nuestros estados mentales, dado que, al parecer, la clase de teorías esbozada hasta aquí no cuenta con ella nada

más que de una forma "lateral" y "secundaria"—como recurso heurístico, como se comentó anteriormente—, si es que no la hace desaparecer definitivamente. Si la TRM ha conseguido ofrecer una teoría válida y coherente acerca de las actitudes proposicionales y de su naturaleza, entonces el sentido real que pueda tener decir que las actitudes proposicionales son estados subjetivos conscientes parece ser prácticamente nulo. Desde luego, los problemas peores para la fundamentación de la psicología científica empiezan a partir de aquí, pues ¿qué es de la presencia de mí mismo ante mí mismo, como decía Merleau-Ponty, en esta clase de consideraciones objetivas del ser psíquico? ¿No habría que fundamentar realmente en esta presencia toda la psicología como ciencia? Al final de todas estas inquisiciones, volveré a considerar mucho más detalladamente que en el capítulo primero qué es lo que la fenomenología husserliana podría decir y aportar a la fundamentación de la psicología contemporánea. En aquel capítulo sólo traté de mostrar que en la propia filosofía fenomenológica se encuentran algunos elementos que son decisivos para su elaboración, pero que, bien pensados y sometidos a cierta crítica, no pueden eludir enfrentarse con el problema de la inteligencia encarnada. Y son precisamente esos elementos los que dan pie a poner en marcha la investigación objetiva del ser psíquico.

SEGUNDA PARTE.

**LA SEMANTICA DE LAS
ACTITUDES PROPOSICIONALES.**

CAPITULO 3.

PERCEPCION, EXISTENCIA, REFERENCIA Y ERROR.

& 1. Prolegómenos.

La mayoría de las reflexiones precedentes han estado dedicadas a la ontología del ser psíquico cuando ésta se investiga desde el punto de vista de la filosofía objetiva, es decir, bajo el supuesto general del realismo científico, entendiendo por realismo científico la defensa de la doble afirmación de que el mundo, especialmente el mundo de los objetos físicos, existe con independencia de los sujetos que lo conocen, y que la mejor manera con la que cuenta la humanidad para conocerlo es la ciencia natural.

Según el hilo de aquellas reflexiones, se expusieron primero las tesis fundamentales del fisicalismo estricto positivista cuyo conjunto ha constituido en nuestro siglo una fundamentación ontológica y epistemológica peculiar y pujante de la ciencia psicológica. El ser psíquico es, en última instancia, ser físico, o, lo que viene a significar lo mismo: no hay ninguna categoría propiamente psicológica.

En oposición abierta con esta clase de fisicalismo, corrientes importantes de la filosofía de la mente contemporánea han argumentado que se puede seguir siendo materialista sin ser reduccionista en relación a las propiedades psicológicas de los organismos. Esta línea de discusión, que aquí

ha sido seguida muy de cerca al hilo, sobre todo, de la obra de Jerry Fodor, arrojó como resultado la defensa, a la par, del monismo materialista de las sustancias y del dualismo de las propiedades psicológicas y físicas. En general, los "seis dogmas del ser psíquico" expresan de una manera condensada las tesis más básicas que un gran número de autores actuales están dispuestos a defender--a pesar de sus divergencias en otros puntos específicos importantes--en relación a la ontología peculiar e inestable del ser psíquico.

Finalmente, se esbozaron los compromisos ontológicos de la psicología cognitiva, tal y como suele ser típico exponerlos, así como la manera en que se podría poner en armonía estos compromisos con los dogmas del ser psíquico, los cuales están a la base de aquellos compromisos cumpliendo el papel de fundamentación ontológica.

Todo esto se hizo con la mira puesta en la consecución de un primer cuadro unitario en relación a las cuestiones ontológicas que están en el fondo de la ciencia psicológica. La idea de asimilar los organismos a los computadores provee, fuera de toda discusión, de un marco de referencia inevitable en este punto. La consecuencia última y más prominente de las indagaciones previas fue la de que tiene que haber representaciones o "símbolos" mentales, símbolos con propiedades semánticas y sintácticas características, pero símbolos que, en virtud de su superveniencia con los estados físicos del sistema nervioso--o, incluso, con los estados físicos de otros sistemas materiales complejos--, tienen también capacidad causal. Precisamente, el difícil problema de la causalidad de los estados mentales, cuando se plantea en el marco de una metafísica materialista, fue el que más páginas consumió en el capítulo previo. En este que aquí se inaugura el interés, sin embargo, se escora más bien hacia la investigación de las propiedades semánticas y sintácticas de estos supuestos símbolos mentales de naturaleza últimamente material: hacia su

interpretación. Si, como se expresaba de una manera tosca, una representación mental es un estado físico complejo que significa, o en giro más exacto, que está sujeto a evaluación semántica, y los contenidos y características intencionales y semánticos de las actitudes proposicionales son los que son porque los enunciados del lenguaje del pensamiento son los que son--las representaciones mentales, decía Fodor, son los objetos inmediatos de las actitudes proposicionales--, entonces interesa saber algo más acerca de estas representaciones mentales. Entre otras cosas, por ejemplo, interesa saber si ellas mismas tienen estructura, al igual que reconocer si la tienen también sus objetos.

La manera de abordar este nuevo ámbito de problemas es acudiendo, como resulta irremediable y canónico desde hace tiempo, a la teoría semántica de las actitudes proposicionales y, en concreto, a las preguntas que inquieren por las características lógicas de las expresiones que usamos para adscribir actitudes proposicionales a otros sujetos. Si se observa despacio, esta forma de proceder general en relación a la investigación de las propiedades semánticas y estructurales de las actitudes proposicionales, es el correlato preciso, en su estilo general, de la estrategia que se adoptó cuando se quiso abordar el estudio de las propiedades ontológicas del ser psíquico. En uno y otro caso se adopta normalmente el punto de vista heterofenomenológico, el de la tercera persona. Sin embargo, si en esta última ocasión el instrumento heurístico para profundizar en la ontología propia y especial del ser psíquico fue el método de la ciencia natural, en el caso de la investigación que ahora se propone lo será las intuiciones lingüísticas, a veces no demasiado claras, que nos acompañan en todo momento cuando realizamos adscripciones de actitudes proposicionales a otros sujetos.

& 2. Yo veo, tú ves, él ve. Yo creo, tú crees, él cree. Materiales para una teoría intersubjetiva del conocimiento.

"En esto, descubrieron treinta o cuarenta molinos de viento que hay en aquel campo, y así como don Quijote los vio, dijo a su escudero:

- La ventura va guiando nuestras cosas mejor de lo que acertáramos a desear; porque ves allí, amigo Sancho Panza, donde se descubren treinta, o pocos más, desaforados gigantes, con quien pienso hacer batalla y quitarles a todos las vidas, con cuyos despojos comenzaremos a enriquecer; que ésta es buena guerra, y es gran servicio de Dios quitar tan mala simiente de sobre la faz de la tierra.

- ¿Qué gigantes? -dijo Sancho Panza.

- Aquellos que allí ves -respondió su amo- de los brazos largos, que los suelen tener algunos de casi dos leguas.

- Mire vuestra merced -respondió Sancho- que aquellos que allí se parecen no son gigantes, sino molinos de viento, y lo que en ellos parecen brazos son las aspas, que, volteadas del viento, hacen andar la piedra del molino.

- Bien parece -respondió don Quijote- que no estás cursado en esto de las aventuras: ellos son gigantes...

Y diciendo esto, dio de espuelas a su caballo Rocinante, sin atender a las voces que su escudero Sancho le daba, advirtiéndole que, sin duda alguna, eran molinos de viento, y no gigantes, aquellos que iba a acometer. Pero él iba tan puesto en que eran gigantes, que ni oía la voces de su escudero Sancho, ni echaba de ver, aunque estaba ya bien cerca, lo que eran...". (Don Quijote de la Mancha. Parte Primera. Capítulo VIII)

En este breve texto aparecen condensados de una manera literaria casi todos los problemas que pueden aparecer en relación a la percepción, a la existencia de los objetos físicos que pueblan el mundo natural, a la referencia de los términos singulares, a la adscripción de actitudes proposicionales a otros sujetos, y, finalmente, a la explicación de la conducta de esos sujetos cuando dicha conducta se explica desde los supuestos de la psicología popular.

Suponemos que don Quijote veía frente a sí un gran número de fieros gigantes. Incluso creía poder señalarlos claramente ("porque ves

allí..."). Y como él los veía, pensaba que los podía ver también su escudero Sancho; que éste compartía el mundo de los objetos que él experimentaba visualmente; que este mundo era, en definitiva, experimentable de una manera intersubjetiva. Pero Sancho, cuya cabeza no estaba en ese momento tan llena de fantasías caballerescas, sólo veía molinos de viento donde su señor descubría gigantes en actitud combativa y amenazadora. Así se lo hizo saber. Por otra parte, don Quijote, sobre el fundamento de su experiencia visual subjetiva, o de algún elemento peculiar de ella, creía que allí *había* gigantes. Como sucede casi constantemente en la vida cotidiana, "infirió" lo que hay en el mundo a partir de lo que vio, y nosotros, en virtud de lo que don Quijote dice y hace sobre la base de su experiencia visual subjetiva, le adscribimos la creencia *de dicto*: don Quijote cree que allí hay gigantes. Sancho Panza, con la misma especie de fundamento, creía que allí sólo *hay* vulgares molinos de viento, y no gigantes.

Algunos temas de gran dificultad se anuncian ya aquí tímidamente. El primero de ellos es saber cómo es posible que dos sujetos delante del mismo objeto "vean" dos cosas diferentes. El segundo es saber en qué consiste la diferencia entre "el mundo de don Quijote" y "el mundo de los objetos mundanos". Que el "mundo" de don Quijote no coincide exactamente con el mundo de los objetos físicos mundanos experimentables intersubjetivamente parece desprenderse fácilmente de la narración de Cervantes. El "mundo" de don Quijote está poblado por gigantes. El mundo de los objetos físicos por molinos. El tercer problema es el de la referencia. El uso pretendidamente referencial que don Quijote hace de muchas de las expresiones con las que cuenta de antemano por ser un hablante competente para expresar el contenido de sus diferentes estados mentales está en una conexión estrecha muy particular con su creencia de que *existen* los objetos a los que él tiene la intención de referirse.

Aparte de estos problemas, parece intuitivamente claro que lo que caballero y escudero creían acerca de lo que veían explica de una manera simple su conducta. Don Quijote creía que allí había gigantes y que "quitar tan mala simiente" era la acción digna que todo caballero andante debía acometer. Por eso espolea su enjuto caballo para dirigirse, lanza en ristre, hacia donde se encuentran "los gigantes". Sancho, por el contrario, permanece en actitud pasiva gritando a su señor para que desista de tan descabellada empresa. Así, pues, los juicios de existencia que los sujetos están dispuestos a sostener y que formulan a partir de "inferencias" que tienen a su base experiencias sensibles visuales, o de otro tipo, de objetos y estados de cosas, son esenciales para cualquier teoría que quiera explicar la conducta de esos sujetos. Sin embargo, así lo dice Cervantes y lo cree Sancho, don Quijote estaba profundamente *equivocado*. Lo que parecían gigantes eran molinos. Don Quijote y Sancho disentían acerca de cuáles eran los juicios de existencia que en esa situación particular estaba realmente justificado sostener.

Para adelantar un tema que se tratará más pormenorizadamente y que llegará a alcanzar gran importancia, diré que esta cuestión, la cuestión de hasta qué punto se está justificado para tener por verdaderos o por falsos ciertos juicios existenciales fundados sobre experiencias sensibles particulares, parece que sólo se puede decidir sobre la base de acciones, experiencias y sucesos futuras, pasadas y presentes de índole muy variada. De este modo, por ejemplo, la historia dice que las aspas de uno de los molinos lanzaron por los aires a don Quijote y que éste cayó al suelo muy maltrecho.

-Válame Dios!--dijo Sancho--. ¿No le dije yo a vuestra merced que mirase bien lo que hacía, que no eran sino molinos de viento, y que no lo podía ignorar sino quien llevase otros tales en la cabeza?

Pero, según parece ser, es posible dudar de, y negar incluso, esta clase de hechos y "evidencias". Se puede tener realmente la cabeza llena de molinos. Parafraseando a Fodor, no basta con estar delante de un molino de viento para verlo *como* un molino de viento (Fodor 1987, 117. Véanse ejemplos semejantes en Castañeda 1977, 340; Goldman 1977, 264-265 y Smith 1984, 68-69). Para concebir y justificar esta posibilidad es preciso suponer algunas condiciones que paso a considerar.

& 3. El transfondo cognitivo y la causación sensorial.

Un ejemplo similar al de don Quijote servirá para analizar y destacar la primera de estas condiciones. Imaginemos a una persona de nuestro entorno cultural que mira por primera vez al conocido dibujo que puede representar tanto a una vieja como a una mujer joven y elegante. Es un hecho que en estas condiciones relativamente ambiguas sea habitual reconocer o bien sólo a la vieja o bien sólo a la mujer joven, pero no a las dos. Supongamos que en la primera ojeada se identifica primero a la mujer joven. La pregunta es: ¿cómo es posible estar delante de un dibujo que representa también a una vieja y no verlo como un dibujo que representa a una vieja? Esta pregunta es, en algunos aspectos, paralela a esta otra: ¿cómo es posible estar delante de un molino de viento y no verlo como un molino de viento?

Si nos atenemos todavía al ejemplo del dibujo se puede lograr que la persona llegue a ver a la vieja por el simple procedimiento de ir indicándole dónde se encontrarían distintas partes de la fisonomía de la anciana. Si se le señala sobre el papel dónde está la nariz, dónde la boca, dónde la barbilla, dónde el pelo, etcétera, será capaz de advertir de pronto que el dibujo representa también a una vieja. Para dar cuenta de este hecho pensemos que los nuevos datos con los que se va proveyendo al

sujeto cuando se le indica dónde se localizan estas partes del rostro de la vieja se caracterizan provisionalmente como *creencias* acerca de la forma de cierta nariz y de su posición en el dibujo, acerca de la forma de cierta barbilla y de su posición en el dibujo, etc. Llamaré al conjunto de estas creencias "el transfondo cognitivo" del sujeto (cognitive background), o parte de él, tal y como es uso canónico hacer.

Por consiguiente y según este ejemplo, una condición necesaria para estar delante de un dibujo que representa también a una vieja y poder verlo como un dibujo que representa a una vieja es que el sujeto tenga alojadas en su transfondo cognitivo cierto número de creencias, relacionadas de una manera holista. Este transfondo cognitivo debe influir decisivamente, si es que no causalmente, para que aparezca súbitamente en el sujeto la creencia consciente de que el dibujo representa también a una vieja. Una vez en posesión de estas creencias en su transfondo cognitivo el sujeto podrá ponerlo en juego para cambiar a voluntad su reconocimiento alternativo de la mujer joven y de la vieja. Así, pues, el transfondo cognitivo debe *mediar* de una manera esencial y compleja lo que vemos y lo que creemos acerca de lo que vemos así como, en general, el contenido de todo acto intencional individual (Mohanty 1984, 24; Searle 1983, 54-55).

Aunque con estas aclaraciones se haya podido arrojar alguna luz sobre las condiciones que permiten a un sujeto ver algo como siendo algo—en general un X como siendo Y—, no son en absoluto suficientes. Para seguir avanzando en este problema planteo una pregunta nueva. ¿Sobre qué se "aplica" o "proyecta" causalmente el transfondo cognitivo que permite ver algo como siendo algo? La respuesta que parece ser más obvia es que el transfondo cognitivo se proyecta sobre ciertos rasgos del dibujo—rasgos espaciales y de color—que el sujeto está en condiciones de ver. En aras de la exactitud esta contestación podría desmenuzarse en dos partes. La primera de ellas es que en el dibujo mismo debe estar contenida ya de alguna

manera la posibilidad objetiva de que el sujeto lo vea como representando a una vieja. Si esta posibilidad fuera nula, cero, el transfondo cognitivo no podría colaborar para ver el dibujo como representando también a una vieja. No conozco a nadie que haya sido capaz de ver en ese dibujo una locomotora del siglo XIX con trece vagones. La segunda parte incide en el hecho obvio de que el sujeto debe estar en condiciones de ver el dibujo. Si el sistema visual de la persona que dirige sus ojos al dibujo funcionara especialmente mal, si fuera muy miope y no llevara gafas, por ejemplo, entonces no podría reconocer a la vieja y, evidentemente, tampoco a la joven sea cual sea, por otra parte, su transfondo cognitivo. Que el sujeto humano vea el dibujo como representando a una vieja depende empíricamente del funcionamiento relativamente bueno de su sistema visual.

Por lo demás no es suficiente con que en el dibujo esté contenida la posibilidad de verlo como representando también a una vieja y con que el sujeto esté en condiciones de ver el dibujo con cierta nitidez. Para que se pueda aplicar el transfondo cognitivo, el sujeto humano debe estar frente al dibujo en ciertas condiciones ambientales. El dibujo debe estar bien iluminado, no haber ninguna pared que impida su visión, etcétera. Es un hecho que si no se dieran estas condiciones empíricas adecuadas, entonces el sujeto no podría ver el dibujo como un X que representa Y por mucho que se den las otras tres condiciones mencionadas. En otras palabras, parece que debe haber una cadena causal, no interrumpida irremediabilmente, entre la presencia del dibujo ahí y la experiencia visual del sujeto de ese dibujo. De esa cadena causal debe formar parte fundamental el sistema visual del sujeto humano.

Así, pues, todas estas condiciones son de entrada necesarias para que una persona vea el dibujo como representando también a una vieja: (i) un transfondo cognitivo suficientemente rico y organizado, (ii) la posibilidad objetiva de que el dibujo represente a una vieja, (iii) la posibilidad subjetiva

de que el sujeto vea el dibujo con cierta nitidez, (iv) que se den ciertas condiciones ambientales adecuadas.

Cuando se tiene ante la consideración teórica estos elementos, no resulta inusual ponerlos en juego y presentar la acción del transfondo cognitivo sobre lo que el sujeto ve de esta manera: lo que el sujeto ve es ambiguo en su interpretación y admite, por lo tanto, diversas descripciones, aunque haya descripciones absolutamente impropias. El transfondo cognitivo debe de determinar decisivamente estas interpretaciones. La persona que mira por primera vez al dibujo de la vieja/joven ve evidentemente algo que describe o interpreta en principio como una joven elegante. Pero eso mismo que ve, a lo que llamaré el contenido de su experiencia visual, admite otra descripción. El dibujo representa también a una vieja. Aquí radica la relativa ambigüedad de lo que ve respecto de su posible interpretación, aunque el contenido de su experiencia visual—el campo de formas y colores que él ve—sea idéntico en ambos casos. Pero cuando se le señala dónde está la nariz de la vieja, dónde su barbilla, etc., el sujeto verá también el dibujo como representando a una vieja. Como comentaba antes, esta ambigüedad no puede ser absoluta. Que hay descripciones absolutamente impropias del contenido de la experiencia visual se muestra porque, manteniendo los términos su significado normal y dadas determinadas condiciones, lo que el sujeto ve no podría nunca ser interpretado o descrito como una locomotora del siglo XIX con trece vagones.

Recuérdese que el objetivo de todas estas indicaciones era el de aclarar qué haría falta para que don Quijote estuviera frente a un molino de viento y no lo viera como un molino de viento. Si su sistema visual funciona bien y está orientado adecuadamente hacia el molino de viento y si las condiciones ambientales son adecuadas, etc., entonces, ¿por qué no ve el molino como un molino? Por eliminación, la sugerencia más

inmediata indica que ello se debe a su transfondo cognitivo. Con todo, se podría protestar enérgicamente contra esta manera de presentar las cosas. En el caso del dibujo de la vieja/joven parece plausible achacar tanto al transfondo cognitivo como al hecho de que el dibujo sea limitadamente ambiguo en sus posibles interpretaciones por qué el sujeto falla a la hora de ver también a la vieja. A pesar de todo, el sujeto reconoce la ambigüedad representativa del dibujo, es decir, que el dibujo representa tanto a una joven como a una vieja, en cuanto se le provea con el transfondo cognitivo adecuado. Sin embargo, en el caso de don Quijote la situación no parece en absoluto ambigua. Es claro que lo que allí hay son molinos de viento. Por lo tanto, proveerle con otro transfondo cognitivo, o enriquecer y variar en ciertos aspectos el que ya tiene, no sería de gran ayuda.

Contra esta forma radical de presentar la cuestión piénsese, no obstante, en lo siguiente. Admitamos que don Quijote tiene ciertamente una experiencia visual determinada cuyo contenido está determinado por un campo visual sensible de formas y colores. No hay por qué pensar que en la experiencia visual misma--al menos para un "observador intacto"¹-- esté escrito que ella verse necesariamente acerca de molinos y que lo único que el sujeto hace es "leer" simplemente lo que allí ya está contenido de suyo bien claro y legible. Don Quijote, con todo, y "con todo" quiere decir aquí a pesar de su experiencia visual consciente, dice: aquellos que allí ves son gigantes. Esa es su lectura o descripción de lo que ve, por así decir. Sin embargo, es claro que se le podría preguntar por *otra* descripción de lo que ve de tal forma que, bajo esta nueva descripción, sea posible tener todavía una lectura verdadera del contenido de su experiencia visual y de lo que hay en el mundo (Anscombe 1981, 10). Por ejemplo, don Quijote

¹ Acerca de la noción de observador u organismo intacto véase Fodor (1987, 117).

podría decir que vio algo muy grande que parecía tener también algo muy grande y largo que se movía a sus lados. Esta descripción sería verdadera tanto de un gigante que mueve sus brazos amenazadoramente como de un molino que mueve sus aspas, aunque en el caso de don Quijote estemos dispuestos a admitir que sólo lo es de un molino que mueve sus aspas. De esta manera es como se podría hacer más semejante el ejemplo del dibujo de interpretación relativamente ambigua al caso de don Quijote.

Sea lo que sea de esta cuestión, en el caso de don Quijote hay un problema añadido que no se presenta en el caso del dibujo de la joven/vieja. Mientras que en el primero parece claro que allí no hay gigantes, sino sólo molinos--por eso don Quijote estaba equivocado--, en este segundo "hay" las dos cosas, la joven y la vieja. El sujeto que al comienzo sólo reconoce a la joven no está equivocado respecto de la interpretación del dibujo, aunque sólo reconozca a ésta. Sin embargo, según el relato, don Quijote acepta amarga y resignadamente que *ahora* sí son molinos lo que *antes* eran gigantes, aunque siga sin aceptar que antes estuviera realmente equivocado. Pero, ¿cómo es posible que don Quijote fuera antes incapaz de reconocer la existencia de los molinos y que ahora lo haga sin ningún problema? ¿Cómo se explica su cambio de creencias? Y, sobre todo, ¿cómo es posible que este cambio de creencias no implique el reconocimiento de un error? Empezaré por esta última pregunta, porque es la más fácil de contestar de las tres. Si don Quijote no reconoce su error es porque, presumiblemente, en su trasfondo cognitivo hay otras creencias que le impelen a no retractarse de su antigua opinión. El reconocimiento de un error, por el contrario, supone renunciar, en el curso de la experiencia fenomenológica del mundo, a antiguos juicios y opiniones que se tuvieron una vez por verdaderos. Así, aparte de creer que allí había gigantes, don Quijote también creía que el sabio Frestón le tenía ojeriza y que haría cualquier cosa por fastidiarle la vida. Ahora, después del tremendo batacazo

contra las aspas del molino, estas creencias de don Quijote acerca de Frestón siguen operando hasta el punto de hacerle pensar que sólo este maldito sabio podía haber sido la *causa* de la sorprendente y rápida mutación de los gigantes en molinos, y no que los molinos que ahora ve eran ya desde antes también molinos. Don Quijote no reconoce su error en razón de estas creencias acerca de Frestón, y, a la luz del transfondo cognitivo en que ellas se encuentran, es muy difícil convencerle de él.

Con todo, falta todavía tener en cuenta un elemento especialmente importante para explicar el cambio de creencias de don Quijote. Quizás, incluso, el elemento más importante de todos. Además de su creencia de que el todopoderoso Frestón había hecho una de las suyas, estamos inclinados a pensar que algo esencial ha actuado también para que don Quijote reconozca *ahora* la realidad y existencia de los molinos de viento: los efectos que sobre su cuerpo y sobre Rocinante ha tenido la enorme colisión con las aspas *del molino*. Si bien la sutileza de Cervantes es extraordinaria y no lo diga explícitamente, no es absurdo reconstruir toda la historia de la siguiente manera. Don Quijote, después de todo lo sucedido, dolorido y tumbado sobre el suelo, pensaba que en el momento de arremeter contra los que él creía gigantes, Frestón los había convertido súbitamente en molinos, y que *la causa* de su golpetazo contra el suelo no fue el brazo enorme de un gigante, sino el movimiento de las aspas del molino—como don Quijote diría, de los gigantes ya transformados por Frestón. No hubiera sido deshonroso ni le hubiera quitado la gloria que un gigante le golpeará en singular y desigual batalla, pero sí lo era que le derribara el aspa de un vulgar molino. Este otro elemento fundamental para explicar el *cambio de creencias* de don Quijote—el conjunto total de los sucesos ocurridos a lo largo del tiempo—jugará un papel fundamental más adelante cuando proponga un argumento "pragmático" para establecer la base

última sobre la que descansan nuestros juicios existenciales acerca de objetos sensibles particulares.

Respecto de este tema es importante tener en cuenta las preguntas que se pueden hacer en relación a esos mismos sucesos. Don Quijote, por ejemplo, se preguntaría cómo es que estaba en el suelo tan malherido. Es muy probable suponer que don Quijote no hubiera creído que el poder de Frestón estaba entre medias de todo lo sucedido, si no hubiera caído al suelo brutalmente golpeado por el aspa de un molino. Pero es fundamental suponer también que aunque el aspa no le hubiera golpeado, *tarde o temprano* habría llegado a la misma conclusión. Después de dar lanzadas y más lanzadas contra las paredes y aspas *del molino* habría caído en la cuenta de que *ahora-otro* ahora, claro está-estaba en presencia de molinos y no de gigantes, y que el sabio Frestón había obrado de nuevo este cambio casi milagroso. La tentación de formular un enunciado contrafactual para dar cuenta de esta circunstancia es casi inevitable. Don Quijote no habría creído que la vileza de Frestón era la causa de que estuviera sentado a los pies de un molino, si es que las aspas del molino no le hubieran golpeado.

Nuevas preguntas se encadenan a partir de aquí. ¿A qué se refería entonces don Quijote cuando dijo que allí había gigantes, sobre todo cuando se asume desde un principio que no los había? ¿Qué es propiamente lo que él creía? ¿Cómo se debe explicar sus cambios de creencias? ¿Cómo es que estos cambios, y las creencias mismas, son determinantes de la conducta de los sujetos a los que se les atribuye? Todas estas preguntas necesitarían una respuesta pormenorizada y convincente. Por mi parte soy consciente de que todavía estoy lejos de tener una plausible para todas ellas. El objetivo de este ensayo es meramente el de servir como primera aproximación a alguno de estos problemas.

& 4. Solipsismo, independencia y covariación causal.

Antes de intentar contestar siquiera mínima y tentativamente a estos interrogantes quiero poner de relieve algunos hechos de sobra conocidos acerca de la relación intencional, eje central sobre el que se concentran todas las preguntas previamente planteadas. A la luz de la exposición anterior, una primera característica privativa de la relación intencional respecto de otras clases de relaciones conocidas—de causa, de tamaño, de orden, etcétera—, es que no parece depender de la *existencia* "ahí fuera" en el mundo del objeto con todo su acervo de propiedades reales. Aunque no hubiera gigantes, don Quijote se refería intencionalmente a ellos cuando dijo que allí había gigantes. Los gigantes eran su "objeto intencional". Cuando se empieza a caracterizar así la relación intencional, entonces no cabe duda de que ésta nace con una vocación manifiesta de *solipsismo metodológico*, en el sentido preciso que dio Putnam a este principio. Según dicho principio "ningún estado psicológico presupone la existencia de otro individuo que no sea aquel al que se atribuye el estado" (Putnam 1975, 220). Este principio general, aplicado al caso concreto de don Quijote, diría que los estados psicológicos de tipo visual de este hidalgo y sus correspondientes estados de creencia acerca de existencias que están fundados sobre dichos estados visuales, presuponen esencialmente su existencia como sujeto, no la de los gigantes.

Por otra parte, además de esta primera característica, se puede señalar también que los contenidos de algunos estados intencionales, como los contenidos de ciertos estados de creencia acerca de la existencia de objetos particulares, parecen exhibir un alto grado de *independencia* respecto de su adecuación o inadecuación en relación a los objetos que en la vida cotidiana consideramos normalmente como las causas de las propiedades que tienen nuestros estados mentales de cualidad visual sobre los que se fundan

aquellos estados de creencia. Es como si hubiera siempre un salto entre las "sensaciones" vividas y las creencias acerca del mundo que se fundan sobre esas sensaciones.

Según las consideraciones del párrafo anterior, esta falla habría que poder rellenarla de una forma holista con nuevas creencias y con otros estados psicológicos de naturaleza muy variada, que habrá que especificar en cada caso particular. Así, por ejemplo, el contenido del estado de creencia de don Quijote "allí hay gigantes", en tanto que estado de creencia fundado sobre ciertas experiencias visuales, parece ser independiente respecto de su adecuación o inadecuación a cómo es el mundo y a lo que hay en él. Sin embargo, no es tan independiente respecto de sus otros estados de creencia. En el fondo, es plausible pensar que el grado de independencia que parecen exhibir respecto del "mundo" los contenidos de los estados de creencia que están fundados sobre experiencias sensibles, está en función más o menos directa de todo aquello que se alberga en lo que se llama un tanto sinuosamente "el transfondo cognitivo".

De momento sólo se requiere ahora caer en la cuenta de que para ver un gigante allí donde "sin duda alguna" sólo hay un molino de viento hace falta pensar que la aprehensión de ciertas experiencias visuales como gigantes ha de estar esencialmente mediada por otras creencias. Lo que estoy sugiriendo es, por tanto, que toda aprehensión o es ella misma una creencia o tiene un elemento de *belief* ineludible, creencia que está mediada esencialmente por otras creencias.

El caso límite de independencia es, desde luego, aquel en el que el contenido de la creencia acerca de existencias es completamente independiente de cómo es el mundo y de lo que hay en él. Don Quijote puede creer que Frestón es un mal bicho, creer implícitamente, por tanto, que Frestón existe, creer que está haciendo un uso realmente referencial, *de re*, del nombre propio "Frestón" y que este uso, sin embargo, nunca

sea efectivamente referencial por mucho que él así lo pensara y lo mostrase con su comportamiento efectivo y con su práctica lingüística cotidiana.

Aparte de este rasgo de independencia de las creencias de don Quijote acerca de existencias particulares y según el complicado y sutil texto cervantino, parece destacarse también que nuestro hidalgo no habría pensado que los gigantes movían sus enormes y amenazadores brazos, si el viento no hubiera impulsado las aspas de los molinos. Esta nueva clase de *covariación causal* que parece existir entre las propiedades de los objetos del mundo y los contenidos específicos de las creencias acerca de existencias que se fundan sobre experiencias sensibles particulares, y que se expresa mediante enunciados contrafactuales, ha sido vista por numerosos filósofos a lo largo de la historia de la filosofía, y en especial dentro de la tradición filosófica analítica contemporánea, como esencial para cualquier teoría del conocimiento de los objetos de percepción. Según Goldman, por ejemplo, la tesis de la covariación causal es "una parte integral de nuestro concepto de visión" (1977, 264).

En oposición al principio del solipsismo metodológico, no se puede dejar de señalar aquí que la tesis de la covariación causal presupone la existencia de los dos polos de la relación intencional, al menos para el caso de aquellas experiencias visuales que forman el núcleo central de las percepciones externas. Algo que existe, en un sentido de existencia todavía por especificar con mayor exactitud, debe haber determinado esencialmente el contenido de ciertas creencias acerca de existencias que el sujeto sostiene fundándose sobre sus experiencias, bien visuales o de otro tipo.

Así, pues, y según todo lo que se ha dicho, parece que hay dos clases de covariaciones. En primer lugar está la covariación causal que señalaba hace poco, a saber, aquella en virtud de la cual don Quijote no habría pensado que Frestón era la causa de la súbita transformación de los gigantes en molinos, si no hubiera sido golpeado por las aspas del molino.

Por otra parte está la covariación causal entre las propiedades de los objetos del mundo--las propiedades de los molinos--y los contenidos de las creencias de don Quijote en un determinado instante del tiempo--el contenido de su creencia de que allí hay gigantes. Que don Quijote creyera que allí había gigantes dependería causalmente de que allí hubiera molinos.

Estos tres elementos, (i) el solipsismo metodológico, (ii) la independencia que exhiben los contenidos de las creencias acerca de existencias particulares respecto de los objetos que hay en el mundo y su dependencia holista, sin embargo, respecto de otros estados psicológicos y (iii) la covariación causal entre el acervo de propiedades que tienen los objetos del mundo y los contenidos intencionales específicos de las creencias fundadas sobre experiencias sensibles, han de ser tenidos en cuenta en todo momento por cualquier teoría que quiera dar una explicación completa de lo que es en realidad una percepción externa².

A juzgar por la bibliografía ingente sobre el problema de la percepción, quien quiera que desatienda cualquiera de estos tres aspectos será el blanco seguro de las críticas de algún otro teórico. Y, sin embargo, poner en armonía estos tres aspectos de la teoría de la percepción no es un asunto nada simple. Por ejemplo, cuando se pone en conexión la última reflexión que reclamaba la atención sobre la tesis de la covariación causal e, implícitamente, sobre la tesis de la existencia del mundo, con la primera propiedad de la relación intencional aludida previamente, esto es, con el solipsismo metodológico, entonces en el ejemplo de don Quijote habría

² Mohanty considera, curiosamente, que las dos bestias negras de la teoría de la intencionalidad son, precisamente, la introducción de cualquier teoría causal para explicar o, mejor dicho, para eliminar la intencionalidad como categoría básica, y el holismo (1984, 17). Como espero argumentar, de ninguna de estas teorías tiene que temer la teoría de la intencionalidad. En todo caso, lo que ellas hacen es venir a complicar más, si cabe, las cosas. Pero nadie ha dicho que las cosas sean fáciles.

encerrada una paradoja curiosa. Al menos así resulta ser a primera vista³. La creencia de don Quijote de que estaba enfrentándose con esas viles y cobardes criaturas que llamamos "gigantes", y que éstos estaban moviendo amenazadoramente sus brazos, parece que es independiente, en tanto que creencia con este contenido específico, de cómo es el mundo, supuesto, claro está, que lo hay y que sabemos cómo es. Entender el ejemplo de don Quijote presupone que podemos llegar a saber que el mundo existe, cómo es, al menos en parte, y lo que hay en él. Pero, por otra parte, parece también plausible pensar que don Quijote no habría "visto" ni gigantes ni brazos amenazadores y que, por consiguiente, su creencia no habría tenido el contenido intencional concreto que tuvo, si no hubiera habido enormes molinos manchegos y el viento no hubiera impulsado sus aspas. Con saltamontes a sus pies es de suponer que don Quijote no habría tenido las mismas creencias que tuvo, aunque dado su transfondo cognitivo cabría esperar casi cualquier cosa. Por consiguiente, en este ejemplo concreto, parece que la creencia de don Quijote con su contenido específico no es tan independiente, después de todo, de cómo es el mundo. Al menos en estos casos, hay también una cierta covariación causal entre cómo es el mundo y lo que hay en él, y los estados mentales que un sujeto puede llegar a albergar respecto de ese mundo. En este sentido, el problema consiste en poner en consonancia estas dos exigencias que parecen mutuamente excluyentes: el solipsismo metodológico y la independencia de los contenidos de las creencias acerca de objetos "sensibles"--como cara y cruz de la misma cuestión--y esta supuesta covariación causal.

³ Este dilema ha sido destacado con otras palabras por Robert C. Richardson, 1981, "Internal Representation: Prologue to a Theory of Intentionality." (*Philosophical Topics* 12): 177. Richardson piensa, en oposición manifiesta a Mohanty, que negar cualquiera de estas dos exigencias--la de la covariación causal y la del holismo--supone poner en serio peligro la noción misma de intencionalidad.

Por cierto, que esto que aquí presento como una posible paradoja no ha sido visto por todos los autores como un motivo directo para asumir sin más que una teoría causal de la percepción está en oposición clara con una teoría holista de la interpretación de los contenidos de las experiencias visuales. "Holismo y causa externa son ambas propiedades absolutamente esenciales del contenido y son perfectamente compatibles" comenta, creo que muy sensatamente, Akeel Bilgrami (1987, 206). La cuestión de fondo, según este filósofo, es que "algunas cadenas causales pueden ser más directas que otras, esto es, menos mediadas por un contorno de creencias que otras" (Ibid.), siempre que se dé a las creencias y a otras especies de estados psicológicos, claro está, la posibilidad de ser eslabones en esta cadena causal. La idea básica encerrada en estas dos citas es, pienso yo, que el contenido de las creencias acerca de existencias particulares está determinado tanto por las cadenas causales físicas externas entre las propiedades de los objetos del mundo—los molinos de viento—y el contenido de esas creencias—la creencia de que allí hay gigantes—, como por la cadena holista interna que relaciona entre sí, también causalmente, otras creencias y estados psicológicos con el contenido de la experiencia visual sobre la que se funda la creencia concreta acerca de existencias. Cuanto más grande sea el contorno de creencias y de estados mentales que medie la aparición del contenido de una creencia perceptual, más grande será la posibilidad de independencia de ese contenido concreto respecto del acervo de propiedades de los objetos del mundo.

& 5. Experiencias visuales, percepciones y creencias sobre existencias.

En cualquier caso, y antes de alcanzar un grado adecuado en la investigación que permita abordar estos problemas con el detenimiento que

requieren, ¿qué cabe decir sobre el ejemplo particular de don Quijote? En un principio cabe adoptar dos posturas extremas, como apuntaba antes. La primera de ellas consiste en afirmar que el contenido de la creencia de don Quijote estaba determinado exclusivamente por lo que debemos suponer que veía--los gigantes y sus amenazadores brazos--en unión con sus otras creencias y todo ello con absoluta independencia respecto de los objetos del mundo y de sus propiedades. De este modo se hace del contenido de las creencias fundadas sobre experiencias sensibles una noción puramente "internalista", una noción que sólo tiene sentido cuando, en el análisis de la percepción, nos atenemos exclusivamente a lo que el sujeto es capaz de describir fenomenológicamente sobre sus experiencias sensibles conscientes y al juicio subjetivo fundado sobre esa descripción. En la segunda postura se conviene, por el contrario, que el contenido de dicha creencia, por ser tal creencia una representación errónea, falsa o inexacta (misrepresentation) de cierto estado de cosas en el mundo--la existencia del movimiento de las aspas de los molinos de viento--, es un contenido que tiene realmente por objeto esos mismos molinos de viento, y ello lo sepa o no el sujeto que soporta esa relación intencional. Cuando el objeto de una creencia se caracteriza de esta manera se hablará de su "objeto material" (material object) (Anscombe 1981, 10). Sólo cuando se tiene en cuenta el objeto material es cuando se puede decir cosas tales como que don Quijote vio molinos.

En el primer caso, la tesis del solipsismo metodológico tendría de nuevo un lugar central en la discusión acerca del problema de la intencionalidad. Dentro de la segunda interpretación, el objeto material de la experiencia visual de don Quijote y de su consiguiente creencia basada sobre dicha experiencia, son los molinos de viento, por mucho que él pensara que eran gigantes. Una complicadísima cadena causal que va desde los molinos de viento hasta la aparición de su experiencia visual con su

contenido concreto y de ésta a los estados de creencia fundados sobre ella, pasando por su complejo y múltiple transfondo cognitivo, explicaría idealmente por qué don Quijote, a pesar de todo, estaría refiriéndose a los molinos de viento—aunque él no lo supiera—, y que hablara acerca de ellos cuando dijo que *allí había gigantes*. Por lo demás, si es plausible que el esquema de teoría que hay que desarrollar ha de parecerse en algún sentido a este último que aquí presento, entonces es evidente que algunos elementos en esta cadena causal son objetos que representan o, al menos, objetos que hay que interpretar de esa manera. Son, por ejemplo, creencias alojadas en el transfondo cognitivo. También se supone que la posibilidad de determinar ese objeto material es siempre factible, aunque no, ciertamente, desde el punto de vista de la primera persona. Para poder decir con algún sentido que el objeto material de las creencias de don Quijote son los molinos de viento, por mucho que la adscripción de la creencia "don Quijote cree que allí hay gigantes" esté enteramente justificada a juzgar por lo que él dice y hace, se necesita que el objeto esté determinado desde el punto de vista de la tercera persona, por el punto de vista de la intersubjetividad.

Cuando adscribimos a don Quijote la creencia de que los gigantes mueven sus brazos amenazadoramente, parece haber un lado en el significado de la expresión "don Quijote cree que allí hay gigantes" que alude de una manera única al contenido de su creencia e, indirectamente, a la experiencia visual subjetiva que fundó dicha creencia. Este aspecto subjetivo, al que parece aludir cierta interpretación de las adscripciones de actitudes proposicionales, es al que Smith y McIntyre llaman cautelosamente el lado "experiencial", propiamente fenomenológico, de una actitud proposicional (1982, 4).

Por otra parte, para captar lingüísticamente mejor este aspecto subjetivo podríamos parafrasear la expresión del contenido de su creencia

diciendo que don Quijote *creía percibir* un gigante. Como argumentaré enseguida, la razón para ofrecer esta paráfrasis es que si se quiere distinguir de cierto entre percepciones y alucinaciones y sostener que don Quijote no percibió realmente gigantes, es esencial reservar el uso del verbo "percibir"⁴ sólo para los casos en los que el objeto que se ve es, además, el objeto que, por existir verdaderamente, ha formado parte antecedente de las condiciones causales que han originado que la experiencia visual de ese objeto tenga las características que de hecho tuvo (Searle 1981b, 726). En este caso, vemos ese objeto y lo percibimos. Por el contrario, el sentido de un estado de alucinación es el de ser aquel estado visual cuyas características fenomenológicas *no* han sido causadas en ningún caso por el objeto que se dice estar viendo.

Pienso que esta primera distinción provisional entre estados perceptivos y estados alucinatorios capta de una forma ajustada el uso que hacemos normalmente de esos verbos y el sentido que va implícito en este uso. Es por esta razón por la que no se puede hablar, sin incurrir en ciertos equívocos, de percepciones correctas y engañosas, como hacen a veces algunos autores. Una percepción engañosa no es una percepción, sino una alucinación, una ilusión óptica o un sueño. El motivo por el que se habla en algunas ocasiones de esta manera—como si hubiera dos clases de percepciones—, es que entre las percepciones y las alucinaciones hay un elemento esencial que tienen ambas en común y que es parte suya inseparable, a saber, *un ver un objeto creyendo que existe*. Es este momento visual de ambas clases de experiencias el que originaría hablar de aquella manera equívoca. Sin embargo, que este momento visual dé pie a hablar de una forma realmente justificada de percepciones y de alucinaciones depende de propiedades extrínsecas al estado visual mismo considerado

⁴ Evidentemente, no se trata aquí de legislar acerca de las situaciones en las que se deben usar ciertas palabras, sino de hacer las distinciones pertinentes y fijarlas lingüísticamente.

aisladamente en el tiempo. Estas propiedades extrínsecas deberán ser destacadas por toda investigación exacta y rigurosa de la percepción.

De momento sólo una pequeña indicación. Según las aclaraciones que propongo, contenidas en el sentido mismo de nuestro uso de los verbos "percibir" y "alucinar", estas propiedades extrínsecas dependerían fundamentalmente *de las causas* de las experiencias visuales, sea como sea que haya que especificar y determinar más exactamente estas causas. Según estas razones, se entiende mejor ahora por qué proponía que don Quijote hubiera tenido que decir "me parece percibir un gigante". Una paráfrasis adecuada de este enunciado sería ésta: *creo* firmemente, a juzgar por la experiencia visual que estoy teniendo en este momento, cuya existencia reconozco inmediatamente, y por las propiedades que ella tiene—su vivacidad, su claridad, etc.—que las propiedades fenomenológicas de dicha experiencia han sido causadas por la presencia y existencia efectiva de gigantes allí enfrente. Así, pues, vivir una experiencia visual engendra espontáneamente en el sujeto dos clases diferentes de creencias, aunque íntimamente unidas. Por una parte, la creencia en la existencia de un objeto sensible particular, por otra la creencia en la existencia de una percepción externa, es decir, la creencia de que la existencia de ese objeto ahí enfrente es causa de las propiedades de esa experiencia visual y de su aparición en la corriente de conciencia del sujeto. Llamaré al primer tipo de creencias "creencias acerca de existencias" y al segundo "creencias perceptivas".

El verdadero problema, sin embargo, no es sólo que don Quijote reconozca necesariamente la existencia de cierto estado visual, que él interpreta como ver un gigante, sino que ese estado visual, sea, por así decir, el "disparadero" casi inmediato de un juicio existencial, no acerca del estado visual mismo y de su existencia, sino acerca de lo que hay en el mundo. Parafraseando a Brentano, diré que la percepción interna (*Wahrnehmung*) consiste siempre, cuando menos, en el reconocimiento de

la unión de la nota "existencia" con el correspondiente fenómeno psíquico (Brentano Psychologie 2, 52-53)⁵, de tal forma que toda percepción interna sería siempre un juicio (Urteil) y un conocimiento (Erkenntnis). Don Quijote reconocería inmediatamente como existente su estado subjetivo que consiste en "ver unos gigantes allí". Pero se trata también de que toda percepción externa debe ser contada dentro de la categoría de los juicios o, cuando menos, que toda percepción es un tener por verdadero (ein Fürwahrnehmen), aunque sea erróneo, un objeto (Brentano, Psychologie 2, 51)⁶.

La duda que recoge Brentano, según la cual una percepción es un tener por verdadero que puede ser erróneo, es la que quería introducir precisamente en la paráfrasis que antes ofrecí. Esta paráfrasis, al introducir las expresiones "me parece percibir" o "creo percibir", le permitiría a don Quijote evitar comprometerse de hecho con cualquier juicio de existencia más o menos implícito acerca de si había realmente gigantes ahí enfrente que fueron la causa tanto de las propiedades que de hecho tuvo su experiencia visual como de la aparición de su creencia acerca de existencias fundada sobre aquella experiencia. "Me parece percibir un gigante" significa lo mismo que "tomo por verdadero que allí enfrente hay gigantes" o que "tomo por verdadero que la afirmación de que allí enfrente hay gigantes está bien justificada".

Por lo demás, aunque la teoría de Brentano ofrezca aclaraciones y herramientas terminológicas importantes para el punto que aquí se está

⁵ Véase también, por ejemplo, este texto de Husserl:

En la esencia de un yo puro en general y de una vivencia en general radia la posibilidad ideal de una reflexión que tiene el carácter esencial de una tesis de existencia evidentemente incontrastable. (1950, 109)

⁶ Más recientemente Edward Craig ha propuesto el esbozo de una teoría judicativa de la percepción que sigue muy de cerca esta antigua tesis de Brentano. Véase Craig (1976).

discutiendo, creo que necesita completarse irremediablemente con un elemento nuevo. La investigación de este elemento está motivada por la pregunta acerca del *origen del error* en que se puede incurrir al tomar por verdadero, por existente, un objeto del mundo cuando este juicio de existencia está fundado últimamente sobre una experiencia sensible consciente, bien visual o de otra clase. Creo poder sostener que este elemento necesario no se puede explicar si no se recurre a conexiones causales heterofenomenológicas entre el estado visual mismo y el objeto existente en el mundo que se dice estar viendo. Es porque en la teoría de momento esbozada falta todavía dar cuenta de esta conexión causal por lo que no se debe reducir sin más una percepción externa a un tener por verdadero, porque este tener por verdadero puede ser erróneo, como el propio Brentano reconoce, aunque, eso sí, toda percepción externa incluya obligatoriamente ese momento judicativo o de *belief*. Para tener una percepción externa no basta, pues, con tener por verdadero un objeto, sino que es necesario que ese objeto haya causado realmente el estado visual que pertenece a la percepción como parte inseparable suya y como elemento sensible imprescindible suyo. Por eso, en toda percepción hay un tener por verdadero, es decir, un elemento judicativo y también un elemento causal que pone en conexión la experiencia visual y cierto objeto del mundo. Precisamente el objeto que se dice estar viendo. Ambos aspectos son condiciones necesarias de toda percepción externa. Supongo que en un estado alucinatorio o de enajenación aguda se tienen por verdaderas cosas que no existen y que es precisamente porque no existen estas cosas por lo que se considera que aquel estado es una alucinación y no una percepción. Si el tener por verdadero un objeto—tenerlo por existente—es erróneo, entonces ese tener por verdadero no es una percepción por mucho que el sujeto así lo crea.

Por lo demás, la relación entre la mera experiencia visual consciente y el momento judicativo existencial fundado sobre ella ha sido claramente expuesta por Quinton:

La relación entre la experiencia inmediata y nuestras creencias acerca de las cosas materiales no es lógica, sino causal, una cuestión de hecho psicológica que explica, pero no justifica, aquellas creencias. (Quinton 1973, 193).

(Sobre la justificación de estas creencias acerca de existencias hay todavía mucho que hablar. De este problema me ocupo por extenso en el párrafo 8.)

Según lo que se acaba de decir, una percepción externa no es sólo un juicio existencial fundado sobre una experiencia visual subjetiva consciente y aprehendida de determinada manera—como un X que es Y—, y que es "disparado" psicológicamente, en tanto que juicio, por la existencia de la experiencia visual consciente; sino que contiene un elemento causal ineludible, aunque la existencia de este elemento causal no la pueda determinar siempre con precisión el sujeto que vive la experiencia visual si es que no tiene en cuenta otros estados psicológicos además de aquél en que se encuentra en ese momento. Este elemento causal determina que el juicio existencial fundado sobre la experiencia visual esté realmente justificado. Por lo demás, es claro que la determinación de la justificación del juicio existencial por parte del elemento causal es independiente de la existencia individual de la propia experiencia visual. Insisto en que la justificación de este componente causal únicamente se puede especificar cuando se tienen en cuenta más elementos de juicio que los que ofrecen al sujeto la mera experiencia visual consciente y el momento judicativo existencial fundado sobre ella.

En cualquier caso, aquello sobre lo que inciden estas reflexiones es en que este elemento causal no puede ser parte de un mero análisis descriptivo del contenido fenomenológico o intencional de un estado visual cuando éste se considera aisladamente en el tiempo⁷. Otra cosa bien distinta es afirmar que la experiencia visual no formaría parte de una percepción externa si no tuviera la siguiente condición de satisfacción: hay un gigante allí enfrente y el hecho de que el gigante exista allí causa la experiencia visual sobre la que se funda finalmente el juicio de existencia en que consiste también necesariamente la percepción.

Por esta circunstancia no puedo estar de acuerdo con Searle cuando cree tener base suficiente para defender que "la experiencia de la causación entra en la experiencia visual misma" (1981, 726. También Searle 1983, 74). No, en la experiencia visual tomada aisladamente en el tiempo no hay *experiencia* de la causación⁸. Considérese simplemente la razón que aporta Searle para defender lo contrario. Para este filósofo también las experiencias visuales tienen condiciones de satisfacción. Según Searle, la condición de satisfacción de una experiencia visual se divide en dos elementos: (a) que hay o existe el objeto que se dice estar viendo y (b) que "estas condiciones de satisfacción deben causar la propia experiencia visual cuyas condiciones de satisfacción ellas son" (1981, 728. Véase también Searle 1983, 48). Creo, sin embargo, que la condición (b) así formulada no puede ser cierta para las experiencias visuales en general. Si lo fuera, no habría forma de estar jamás en un estado alucinatorio o ilusorio.

En el fondo, creo que Searle se contradice en la misma página cuando, después de hacer estas afirmaciones, dice un poco más abajo:

⁷ Algo parecido defiende también Smith (1984, 75).

⁸ Locke ya sabía esto cuando escribió el Ensayo (& 25 Libro II, capítulo VIII.)

El hecho de que la causación sea siempre parte del contenido y, así, parte de las condiciones de satisfacción de la experiencia visual, no implica que la condición causal se satisfaga de hecho siempre. (1981, 728).

Pero justamente por eso, porque la condición causal no se satisface necesariamente cuando se tiene una experiencia visual de un objeto o estado de cosas--de un X como siendo Y--, es por lo que esa condición no puede ser parte del contenido de *todas* las experiencias visuales. Si lo fuera, entonces la condición causal sería también parte del contenido de las experiencias visuales que son alucinaciones. Como mucho, la condición casual sería parte del contenido de la creencia perceptiva fundada sobre la experiencia visual, es decir, parte de las condiciones de satisfacción del contenido de un juicio existencial fundado sobre esa experiencia visual.

A propósito del problema de si la causación entra o no entra a formar parte constituyente del contenido de la experiencia visual, Mohanty hace una crítica a Searle que es interesante y, según mi parecer, adecuada en sus líneas generales. Según Mohanty lo que hay que reconocer como una afirmación acertada es que en toda experiencia visual hay un componente de pasividad ineludible, un ser o sentirse afectados respecto de la aparición de la experiencia visual. Ella es vivida por el sujeto como apareciendo involuntariamente en su corriente de conciencia. Sin embargo, lo que no se puede admitir a partir de aquí sin más argumentaciones es que este fenómeno de la pasividad deba ser explicado *siempre* mediante la afirmación de una relación causal entre un objeto existente en el mundo y la experiencia visual de ese objeto que tiene supuestamente el sujeto (Mohanty 1984, 21). No obstante y tal y como he indicado, aunque esté de acuerdo con Mohanty en su oposición a Searle, de aquel filósofo me alejo si es que el reconocimiento de su crítica implicara el reconocimiento de su validez absoluta. Mi tesis, por el contrario, será que la experiencia

fenomenológica *de un sujeto a lo largo del tiempo* sí que da pie a considerar e inferir sin duda que la causalidad es un elemento esencial de las percepciones externas y, precisamente, que es el elemento que determina cuándo una experiencia visual es parte es una percepción externa y cuándo parte de una alucinación.

& 6. Dos sentidos del verbo *ver*: *de re* y *de dicto*.

Cuando se tiene en cuenta este elemento causal que parece ser tan necesario para construir una teoría adecuada de la percepción externa, las cosas se complican. Desde el punto de vista de la intersubjetividad, Cervantes puede decir, sin que resulte por ello una forma de expresarse especialmente forzada, que don Quijote vio unos molinos de viento, aunque creyera que lo que vio eran gigantes. ("En esto, descubrieron treinta o cuarenta molinos de viento que hay en aquel campo, y así como don Quijote *los vio*...".) Esta adscripción de Cervantes--don Quijote vio molinos de viento--hace hincapié sobre todo en el hecho de que aunque don Quijote no tuviera ni la más remota idea de que lo que vio eran molinos de viento--él pensaba que lo que se movía eran gigantes--, sin embargo, lo que él vio fueron molinos. Es decir, los molinos entraron a formar parte esencial antecedente de la génesis causal de su experiencia visual y determinaron el hecho de que esta experiencia fuera interpretada de la manera en que lo fue dando origen, subsecuentemente, a su creencia con su contenido específico: yo creo que allí hay gigantes, después de pasar, obviamente, por su transfondo de creencias⁹.

⁹ Thomson se contrapone abiertamente a la tesis de que esta hipótesis sea viable (1986, 185). Según este autor si estamos dispuestos a hablar de esta manera debemos renunciar a explicar la relación intencional.

Cuando se pone el acento en el uso del verbo "ver" que está haciendo Cervantes en este contexto "intersubjetivo", se está haciendo un uso, en palabras de Smith y McIntyre, "objetivo" y no fenomenológico de este verbo (1984, 4). Mediante este uso Cervantes hace mención principalmente a la relación física-causal que tiene lugar entre un cierto objeto físico llamado "don Quijote" y otros objetos físicos llamados "molinos".

Por lo demás, la utilización por parte de Cervantes del verbo "ver" en este contexto es una utilización *de re* muy fuerte de dicho verbo, aunque, como se puede observar, en absoluto inusual en castellano. Así, habría dos formas posibles de especificar el objeto de la experiencia visual de don Quijote. Una *de re* que subrayaría su uso objetivo, intersubjetivo, no fenomenológico: don Quijote vio molinos de viento. Otra *de dicto* que resaltaría su uso estrictamente fenomenológico: don Quijote vio gigantes.

Según se caracterice *de re* o *de dicto* aquello a lo que se refiere el verbo ver, se podría especificar, correspondientemente, también *de re* o *de dicto* el contenido de las creencias de don Quijote. En la caracterización *de re* del objeto de la experiencia visual, quien hace la adscripción se compromete muy fuertemente con la existencia del objeto que es el origen de la causación de la experiencia visual de don Quijote. Este es el objeto material.

Por otra parte, el uso *de re* del verbo "ver" por parte de Cervantes implica que no hay ningún estado de conciencia en don Quijote que se pueda describir como el tener conciencia de que allí hay molinos. Es claro que don Quijote no puede haber tenido conciencia de lo que Cervantes dice que vio, dado que, de ser así, don Quijote habría dicho algo muy distinto de "allí hay gigantes". Este uso *de re* del verbo "ver" expresa la existencia de una "conciencia-cero" (Castañeda 1977, 340). El sujeto no es consciente de ver lo que el adscribidor dice que ve. Pero, entonces, ¿a qué

se refiere el uso *de re* del verbo "ver" si no es a nada de lo que el sujeto sea consciente? En primer lugar, Cervantes debe de referirse a algo así como que el sistema visual de don Quijote fue activado causalmente por la presencia efectiva de molinos de viento. ¿Alguien tiene otra idea acerca de cuál puede ser la referencia del uso *de re* del verbo "ver", aparte de la que acabo de mencionar? Esta activación sería una condición necesaria para que haya percepción externa de un objeto mundano. La peculiaridad fundamental de este uso *de re* del verbo "ver" es que "es lógicamente posible que alguien vea un objeto que no satisface... ninguna creencia generada visualmente acerca de él" (Goldman 1977, 262). En el caso de don Quijote, su ver *de re*, por así decir, molinos no satisface ninguna creencia acerca de existencias generada posteriormente. Lo que se generó posteriormente cuando vio *de re* molinos fueron creencias acerca de la existencia de gigantes, no de molinos.

Apoyados en este uso *de re* fuerte del verbo "ver" estaríamos justificados para informar lingüísticamente de la creencia de don Quijote que se funda en dicho uso del verbo "ver" así: don Quijote cree que los molinos son gigantes o, en otro giro, don Quijote cree percibir gigantes, pero no percibió ninguno. El segundo caso, aquel en el que se caracteriza *de dicto* la experiencia visual de don Quijote, la creencia *de dicto* correspondiente sería: don Quijote cree que allí hay gigantes.

En cualquier caso, sea lo que sea de estas cuestiones, se adelantan varias otras que son también esenciales dentro de toda esta discusión. En primer lugar, el problema de la referencia de los términos singulares y, en concreto, de los deícticos. Don Quijote dijo que *allí* había gigantes. En segundo lugar y especialmente, el problema de la existencia de los objetos de percepción a los que supuestamente nos referimos cuando usamos aquellos términos de una manera referencial singular.

& 7. Indexicalidad y percepción externa.

Un resumen adecuado de todo cuanto se ha dicho hasta este momento incluiría necesariamente algunos elementos analíticos mínimos. Las experiencias visuales son parte imprescindible tanto de las experiencias que llamamos "percepciones" como de las experiencias que llamamos "alucinaciones". Es cierto que entendemos adecuadamente la diferencia que debería existir entre una percepción y una alucinación. A su vez, una experiencia visual, que es parte que no puede faltar en las percepciones y en las alucinaciones, consta, al menos, de tres elementos insuprimibles: (i) un elemento consciente--la experiencia es siempre experiencia de un sujeto que la tiene y que puede hablar con sentido acerca de ella--, (ii) un elemento sensible--el contenido de la experiencia visual-- y (iii) un elemento predicativo e interpretativo--en el campo visual están presentes ante el sujeto particulares con determinadas propiedades (Castañeda 1977, 295). El elemento sensible aludido es un campo visual y, en general, un campo sensible dado conscientemente al sujeto y que posee, gracias al elemento predicativo que se le acopla, una estructura determinada, posiblemente proposicional, del tipo "esto es esto" o "esto tiene esto"; esto es un gigante con brazos largos, por ejemplo.

Además de este elemento predicativo, en toda experiencia visual hay la posibilidad de destacar un elemento referencial indexical que el sujeto "vive" y puede expresar, y en torno al cual puede haber mayor o menor número de predicaciones, verdaderas o falsas. En cualquier caso, un *allí* o *aquí* hay un X que es Y. Ahora la cuestión es saber si este componente referencial indexical destacable siempre en toda experiencia sensible y que se da en ella necesariamente ante la "mirada" del yo que la vive, se refiere al lugar relativo que ocupa un objeto físico respecto de donde se encuentra el sujeto, y si ese elemento referencial indexical depende, para que sea

realmente referencial, de la existencia real como unidad de ese objeto en el mundo, de tal modo que en la experiencia sensible no habría elemento indexical realmente referencial si no hubiera objeto.

Si la distinción entre indexicalidad realmente referencial e indexicalidad presunta tuviera sentido, entonces sería imprescindible distinguir entre la intención o el querer referirse a un objeto y el referirse a él realmente. Como puede imaginarse fácilmente, esta distinción depende enteramente de que se admita que la causación de la experiencia visual por parte del objeto que se dice estar viendo "allí", es un componente esencial para determinar cuándo una experiencia visual es parte de una percepción externa y cuándo parte de una alucinación. En las percepciones externas, en las que el juicio existencial debe estar justificado, el elemento indexical dado por la mención del yo será realmente referencial. No así en las alucinaciones en las que el juicio existencial no está justificado. Si se admite esta tesis, entonces el uso referencial del adverbio de lugar "allí" que utiliza don Quijote para referirse al supuesto lugar relativo que ocuparían realmente los gigantes respecto de donde él se encuentra, sería pretendidamente referencial y no realmente referencial. Ciertamente, lo que don Quijote dice implica que él contestaría sincera y afirmativamente a la pregunta "¿hay enfrente de usted, justamente "allí", gigantes?". Lo mismo se podría decir de Sancho si se le preguntara respecto de los molinos de viento. En general, desde el momento mismo en que un sujeto crea firmemente que en él se está dando la ocurrencia de una percepción externa--la haya o no realmente--, esto es, desde el momento en que se dé en él una experiencia sensible visual con determinadas propiedades, habrá en su vivencia, como mínimo, un elemento *demonstrativo* o *indexical* inherente, esto es, un "aquí", un "allí", un "éste" o "aquél" que se referirá al objeto o estado de cosas que el sujeto cree estar percibiendo y al cual se quiere referir lingüísticamente. Este elemento demostrativo existente en

toda experiencia visual ha sido especialmente destacado, entre otros, por Castañeda (1977, 295), Husserl (1950, 220), Miller (1984, 45-46), Searle (1981b, 725-726) y Smith (1984, 61-62). La percepción, dice Smith, "es un tipo de conciencia demostrativa única" (1984, 61-62).

Como se ha indicado anteriormente, cualquier ejercicio de análisis de fenomenología sincrónica de las experiencias sensibles, amenizado con la cuestión de si los juicios de existencia que se fundan sobre dichas experiencias son válidos o no, lleva de suyo a considerar la dificultad que se encierra cuando se quiere tener una teoría adecuada de la percepción externa. Este problema está íntimamente entrelazado, dicho sea de paso, con el de si el elemento referencial indexical que el sujeto vive al vivir esa experiencia sensible con su estructura proposicional, es signo de la existencia del objeto mismo. El uso referencial de los demostrativos por parte de los sujetos que los usan para expresar sinceramente el contenido de sus experiencias visuales, es un signo indirecto importantísimo de que el sujeto, siquiera sea implícitamente, lleva a cabo la formulación de estos juicios existenciales. Pero como decía, este uso pretendidamente referencial será realmente referencial cuando la verdad del juicio existencial, del que los indexicales son signo revelador, esté realmente asentada. Ahora bien, parece que la cuestión de la referencialidad de los indexicales depende de la justificación de nuestros juicios existenciales. Luego todo el problema se concentra a partir de aquí en esta pregunta: ¿cómo podemos saber que unos juicios existenciales, los que están insertados en las percepciones, están justificados, mientras que los juicios existenciales que constituyen también las alucinaciones, por el contrario, no lo están? No, como se ha insistido a lo largo de las páginas anteriores, ateniéndose aisladamente a las experiencias sensibles, que son partes que no pueden faltar en las percepciones y en las alucinaciones. Por lo demás, debería resultar claro que hasta este momento sólo se ha presentado *el sentido* que debería diferenciar

las percepciones externas y las alucinaciones y, por tanto, el criterio "nominal" que las distingue. La posibilidad de destacar este criterio nominal remite a la noción de causa de las experiencias visuales y sensibles en general. Lo que no hay todavía es la posibilidad de ofrecer este criterio de una manera bien fundamentada. A partir de aquí, ¿cómo sería posible argumentar para asegurarse de la existencia de estas causas y cómo esta certeza podría colaborar para distinguir las percepciones de las alucinaciones, la indexicalidad real de la presunta?

& 8. Conocimiento por familiaridad y conocimiento por descripción.

El comentario al texto del Quijote citado anteriormente sirvió, entre otras cosas, para resaltar el problema de la referencia cuando se contextualiza en el plano más amplio de la explicación de la percepción externa, de la conducta y de la especificación de los objetos de creencia. ¿A qué se refería don Quijote cuando decía que "allí hay gigantes"? Para desbrozar este interrogante, recordaré algunas tesis interesantes de Russell acerca de los nombres propios, las descripciones definidas y el conocimiento. Advierto desde ahora mismo que me serviré de ellas con alguna liberalidad, porque mi interés no es tanto exponer paso a paso de una manera pormenorizada la teoría de Russell, cuanto el hacer uso de su acervo conceptual para aclarar estas cuestiones donde sea necesario. Sin embargo, creo que tiene un interés añadido recordar estas teorías de Russell. En ellas se contienen tesis que han originado un gran número de teorías semánticas actuales, bien como continuación natural de aquellas otras, bien como reacción en contra. Lo que pondría de manifiesto este recuerdo es lo íntimamente unidas que estaban en su origen algunas cuestiones semánticas básicas con la teoría del conocimiento, aspecto que muchas veces

puede olvidarse con facilidad. Por ejemplo, tarde o temprano muchos autores se han visto obligados a relacionar la teoría de los objetos con los que estamos en una relación cognitiva directa de familiaridad, con las teorías de los nombres propios y las descripciones definidas. Si algún mérito tiene este trabajo es intentar hacer ver más claramente esta relación.

De acuerdo con Russell nuestro conocimiento de los objetos se divide en dos clases bien definidas: conocimiento por familiaridad (knowledge by acquaintance) y conocimiento por descripción (knowledge by description). El conocimiento por familiaridad es la mejor clase de conocimiento que podemos tener de los objetos, y es el fundamento de todas las demás clases posibles de conocimiento (Russell 1989, 26). Esta especie de conocimiento tiene, al menos, estas dos características fundamentales: (i) es infalible y (ii) es completo respecto de su objeto. Se infiere, correlativamente, que el conocimiento por descripción no es ni infalible ni necesariamente completo respecto de su objeto.

Russell afirma que estamos familiarizados con un objeto cuando tenemos "una relación cognitiva directa con ese objeto, esto es, cuando somos directamente conscientes del objeto mismo" (1956, 209 y 1989, 25). Pero, ¿qué significa esta relación cognitiva directa?

Cuando hablo aquí de una relación cognitiva directa no quiero decir la clase de relación que constituye el juicio, sino la clase de relación que constituye la representación (presentation). (Russell 1956, 209)

En otro lugar dice que esta relación cognitiva directa con un objeto equivale a conocer ese objeto sin mediación alguna de procesos de inferencia (1989, 25), y más tarde afirma que "decir que S tiene familiaridad con O es esencialmente la misma cosa que decir que O está presente a S" (1956, 210). A partir de estos textos se deduce que ser conscientes del

objeto mismo quiere decir: tener una presentación de él no mediada inferencialmente, una relación cognitiva privilegiada.

Ateniéndose a las características mencionadas, ¿de qué clase de objetos cabe tener conocimiento por familiaridad? Muchas veces comenzar por los análisis de los ejemplos más triviales e inmediatos resulta fructífero. Básicamente, consideramos normalmente que hay dos formas posibles de entender cómo somos conscientes del objeto mismo, esto es, cómo nos representamos un objeto, cómo lo tenemos presente. Intuitivamente, cuando veo una manzana que está encima de la mesa, digo, con algún refinamiento técnico, que me represento perceptualmente la manzana. ¿Quiere decir esto que estoy familiarizado con la manzana en tanto que objeto físico? Más adelante se verá.

En segundo lugar, si un amigo me dice "hay un monstruo que vive debajo de la superficie helada del lago Mendota" y yo entiendo el significado de sus palabras, entonces me represento ciertamente algo. Mediante la comprensión del significado de sus palabras me hago consciente de algún estado de cosas posible que, caso de existir, haría verdadero el enunciado de mi amigo. En este segundo caso, parece que estoy familiarizado no sólo con el significado de su oración, sino también con este estado de cosas posible, en tanto que posible. Si estoy familiarizado con el significado de sus palabras, si entiendo lo que dice, no puedo evitar, por así decir, representarme el estado de cosas posible que, caso de existir, haría verdadero o cumpliría su enunciado: el estado de cosas que constituiría la condición de satisfacción de aquella descripción definida. Por consiguiente, cuando estamos familiarizados con el significado de las palabras y de los enunciados, cuando entendemos sus significados, tenemos conocimiento infalible y completo no sólo de estos significados—cualquier "cosa" que ellos sean—, sino también de ciertos estados de cosas posibles que, caso de existir, harían verdaderos los enunciados con significado.

& 9. La existencia y el problema del conocimiento.

Yo digo ahora estar viendo/percibiendo una manzana que está allí encima de la mesa. Pero puedo estar soñando o en un estado alucinatorio más o menos transitorio. En tales ocasiones, sin embargo, podría seguir sosteniendo pertinazmente que veo/percibo una manzana encima de la mesa, aunque no haya una tal que, por ejemplo, pueda coger y morder. Como se comentó anteriormente, por lo que se refiere exclusivamente a las características fenomenológicas sincrónicas de la experiencia subjetiva que describo como la visión/percepción de una manzana, no parece haber diferencias intrínsecas esenciales entre percibir una manzana (existente) o meramente "alucinarla", cuando tomo aisladamente uno de estos estados visuales como el único fundamento para concluir afirmativamente que la manzana existe. En ambos casos, la manzana puede presentármese "ella misma en persona"--como en ocasiones dice metafóricamente la fenomenología--con la misma vivacidad. Desde luego, mientras que un sujeto se atenga *de una forma aislada* a la certeza con que se le presenta cualquier clase de experiencia subjetiva individual de carácter "visual" o de otro tipo--sea una percepción, sea una alucinación--para determinar la validez del juicio de existencia acerca del correspondiente objeto en el mundo extralingüístico, no hay manera alguna de que pueda empezar a distinguir entre una percepción y una alucinación, es decir no puede empezar a tener más elementos de juicio, alguno de ellos esenciales, para determinar si la manzana existe o no existe verdaderamente y si ha sido ella la causa de las propiedades de su experiencia visual. Como dice Searle, "del enunciado de que la experiencia sucede no se sigue que el objeto exista" (1983, 75). En este sentido, el contenido de una experiencia visual "no entraña lógicamente una causa particular. La experiencia, después de todo, podría ser una alucinación causada por una droga.... No podemos, simplemente a

partir de la experiencia misma, determinar su veracidad (veridicality)"
(Thompson 1986, 92).

Entre estos elementos se encontraría, quizá, morderla para quitarse el hambre. Desde luego se puede soñar también que se muerde una manzana, pero mi tesis será que este proceso de ir incorporando cada vez más y más experiencias y acciones al mundo del sueño y la ilusión para ensanchar hasta el máximo posible los límites de nuestro escepticismo en relación a la existencia de los supuestos objetos de percepción--a la existencia del mundo en resumidas cuentas--, no puede continuarse *indefinidamente* si se quiere llegar siquiera a plantear el problema del conocimiento, al menos cuando éste se encara al estilo cartesiano: ¿cómo sé yo que existe un mundo de objetos trascendentes independiente de las ideas de todo tipo que encuentro en mí? La determinación de la existencia de los objetos de percepción tiene mucho que ver, así argumentaré, con el conjunto y entrelazamiento variadísimo y coherente de múltiples tipos de experiencias y sucesos que están relacionados con el sujeto psicofísico, considerado como un todo, a lo largo del tiempo. Es más, existencia natural quiere decir justamente aquello mismo que permite que yo pueda determinar la existencia de los objetos de percepción de esta manera diacrónica, histórica. Nada más que eso. Cualquier otra noción que se quiera aplicar para dar sentido a la expresión "existencia de los objetos de percepción" no está justificada. Pero esto es suficiente. No hay que pedir más, pero tampoco menos. Se necesita, en otras palabras, experiencia para establecer estas existencias particulares. Por este motivo creo que el único fundamento con el que podemos contar para decidir, sobre bases racionales, la existencia particular de cada uno de los objetos de los que tenemos experiencias visuales, u otra clase de experiencias sensibles, es esta coherencia de múltiples clases de experiencias--por lo tanto, no sólo visuales-- y sucesos variadísimos a lo largo del tiempo de nuestra vida.

Por lo demás, es evidente que podemos estar sujetos a cometer errores al establecer la existencia de las cosas particulares en cada ahora del tiempo, si lo aislamos del resto de *ahoras* y de lo que sabemos y experimentamos en ellos¹⁰. Sin embargo, que, en general, ha habido y hay objetos de percepción externa--no este o aquel objeto en particular--, puede establecerse con certeza desde el momento en que se haya tenido alguna vez experiencia del error y se entienda adecuada y seriamente en qué consiste esta peculiar experiencia cognoscitiva, posiblemente la más fundamental de todas. Esta experiencia se nos aparece en la conciencia como dada absolutamente en su sentido. Pero la comprensión de este sentido basta para el argumento que quiero presentar. La posibilidad de dudar de la existencia de las cosas mundanas presupone que la experiencia del error se me da y que se me da de forma indubitable, al menos en su sentido originario. En otras palabras, que la experiencia del error no puede ser afectada por la duda misma, porque ésta presupone aquélla. En efecto, Descartes supuso motivo suficiente de duda el tener el más mínimo motivo de duda. De todo lo que se le daba por los sentidos y había aprendido a través de ellos podía dudar, porque había experimentado que *a veces* tales sentidos le engañaban. El hecho de que tuviera ya experiencia del error y entendiera el sentido de esta experiencia posibilitó su capacidad de dudar. La idea del argumento que voy a presentar es que para experimentar *a veces* que los sentidos nos engañan y poder decir eso con sentido, para tener experiencia del error y comprender, en definitiva, la posibilidad de esta experiencia y de la duda misma, se necesita haber estado ya alguna vez en *contacto* con la verdad, esto es, con lo que el mundo es y con lo que hay en él o, al menos, con alguna parte suya realmente existente. Empezar con la reflexión de la duda está muy bien, pero no hay que ir "hacia

¹⁰ Para esta cuestión, véase el final de la Sexta Meditación (Descartes 1967, 235).

adelante" y extenderla todo lo que se pueda en sus límites, sino más bien "hacia atrás". Es imposible incorporar dentro del saco de la duda *todo* lo que se nos da por los sentidos, porque la posibilidad de dudar acerca de la existencia de los objetos sensibles particulares supone que no me puedo engañar acerca de la existencia de *todos* los objetos que se me dan, o se me han dado, por los sentidos.

Las premisas del argumento que a continuación voy a presentar son, por consiguiente, éstas: (i) que hay o tengo experiencia del error o, como mínimo, del sentido de la experiencia del error, (ii) que esta experiencia, como toda vivencia que pertenece a la corriente de vivencias de una conciencia, se me da absolutamente en su existencia y en su sentido¹¹, (iii) que sin esta experiencia el problema del conocimiento, al menos en las versiones cartesianas, ni siquiera puede comenzar a plantearse. La conclusión del argumento será que no puedo dudar *siempre* de la existencia del mundo desde el momento en que quiera tener una explicación de esa experiencia cognoscitiva peculiar. Pero si no hay explicación del error, entonces no hay posibilidad de dudar. El argumento, pues, parte de que la experiencia del error se me da absolutamente en su sentido y se pregunta cómo es posible que se dé esta experiencia, dada ella misma absolutamente.

& 10. Imposibilidad de la duda absoluta en relación a la existencia de los objetos de percepción.

Viajo un día muy caluroso por una de esas rectas interminables que hay en La Mancha y veo un charco de agua sobre el asfalto de la carretera, justamente allí, a unos cuantos metros delante del coche en que voy cómodamente sentado. Pero cuando llego "allí" me doy cuenta de que

¹¹ Esta tesis no la argumentaré aquí, pero véase, por ejemplo, Husserl (1950, 108-109).

no hay charco alguno. Creo eso. ¿Por qué? Porque los charcos mojan los bajos de los coches; se ven mejor cuando se está más cerca de ellos; las llantas de los coches hacen un ruido peculiar, que puedo oír, cuando ruedan sobre el asfalto mojado y despiden también con fuerza el agua por los laterales, etc. Sobre la base de todas estas experiencias y sucesos esperados y que no se cumplen, digo que aquí y ahora—cuando he llegado *allí*—no hay ningún charco. Bien, pero, ¿cómo puedo establecer que *antes* tampoco había charco, es decir, que antes estaba equivocado y que ahora, por lo tanto, no sólo sé cómo es el mundo en este momento, sino también cómo era antes? ¿No podría haberme pasado como a don Quijote? ¿No podría haber sucedido que cuando llego "allí" algo o alguien haya hecho que las cosas cambien de pronto? ¿No podría ser que el mismo calor de la carretera haya evaporado el agua del charco? Admito que eso podría haber ocurrido perfectamente. Sin embargo, es fundamental darse cuenta de que en tal caso *no* hubiera tenido experiencia del error y que es esta experiencia la que se me da con evidencia en su sentido y la que quiero describir en sus propiedades esenciales. Hubiera dicho, como dijo don Quijote, que *antes* las cosas eran de una manera y que *ahora* son de otra diferente, pero no que *me he equivocado*, es decir, que antes no había charco cuando juzgué que sí que lo había. Si el sabio Frestón tiene poder suficiente para hacer que los gigantes se conviertan en molinos, de seguro que no le faltará poder para hacer desaparecer charcos de agua manchegos.

Lo que pone de manifiesto este ejemplo, una vez más, es que el halo de experiencias sensibles y sucesos particulares de toda índole que aquí y ahora puedo tener en cuenta, en virtud de la experiencia previa, para establecer la existencia o inexistencia de un objeto singular no es suficiente para determinar con evidencia la exactitud de cualquier juicio de existencia que haga sobre él en este momento presente. Pero este argumento no está orientado exactamente a la demostración de la existencia de este o aquel

objeto en particular, como ya se comentó, sino a la demostración de que *alguna vez* ha tenido que haber objetos realmente percibidos por mí, es decir, que yo he tenido que estar realmente alguna vez en contacto epistemológico privilegiado con un trozo existente del mundo. Estas afirmaciones no quieren decir por ahora otra cosa más que esto: alguna vez los predicados que son verdaderos "del objeto intencional de <la> experiencia visual son también verdad de la cosa percibida en cuestión" (Hall 1977, 383). En resumen, con el halo de experiencias y sucesos actualmente a mi disposición no he demostrado que antes estaba equivocado cuando dije que existía un charco de agua allí delante, pero tampoco he conseguido explicar la experiencia del error, porque no admito haber cometido ninguno. (Uno mismo hace de *alter ego* de sí mismo a lo largo del tiempo. El paso del tiempo le hace ser a uno su propia intersubjetividad.)

Sigo conduciendo el coche por la misma interminable recta manchega y, al rato, veo de nuevo otro charco de agua un poco más allá. Extremo las precauciones en la conducción. Este parece un charco todavía más grande que el anterior y tengo experiencia de que los charcos grandes varían peligrosamente la dirección del coche. Como no quiero tener ningún problema, cojo el volante con fuerza. De nuevo, llego allí donde está para mí el charco de agua y no veo ninguno. Nada hay tampoco que me indique que aquí y ahora existe uno. El razonamiento a partir de este momento continuaría como en el caso del "charco" de hace unos kilómetros. O bien el sabio Frestón en forma de calor insoportable (el coche no tiene aire acondicionado), ha hecho que el charco se evaporara, o yo estaba equivocado y no había antes charco de agua alguno cuando juzgué que sí que lo había. Alárguese lo que se quiera las experiencias antes de llegar a esta dicotomía, ella siempre se planteará cuando se quiera explicar la experiencia del error. La dicotomía consiste en lo siguiente. Si

me decido por la primera opción, es decir, si creo absolutamente en el poder de Frestón, sigo sin haber explicado y descrito en qué consiste la experiencia del error, o, si se quiere, sin haber dado sus condiciones de posibilidad. Seguiría hablando de que el mundo era antes así y así y ahora de esta y aquella manera. En resumen, no podría distinguir nunca el sueño de la vigilia. Si me inclino por la segunda alternativa, por aquella que dice que yo estaba antes equivocado, he explicado la experiencia del error y he demostrado ahora que no existía un charco de agua cuando juzgué que sí que existía. Esta segunda opción supone, al menos, que ahora estoy en contacto con la verdad, es decir, que ahora el mundo no es simplemente *mi* mundo sino *el* mundo; que mi experiencia visual es, además, el constituyente de una percepción del mundo extralingüístico tal cual él es ahora. También sé ahora cómo era el mundo antes; corrijo mis antiguas opiniones.

Es desde esta perspectiva como se comprende mejor que la experiencia fenomenológica a lo largo del tiempo nos da pie a considerar e inferir sin asomo de duda que la causalidad es un elemento esencial extrafenomenológico de las percepciones externas. Pero este elemento sólo se puede establecer a lo largo de la experiencia fenomenológica del mundo. La experiencia del error se da ella misma en la experiencia fenomenológica a lo largo del tiempo. Me encuentro con ella y con su sentido. Pero el error hace que, cuando quiero explicarlo, dar sus condiciones de posibilidad—que son las condiciones de posibilidad de una experiencia dada absolutamente en su sentido—concluya que alguna vez "mi mundo" es también el mundo de los objetos físicos mundanos. Es así como puedo decir que antes mi experiencia visual no pudo ser causada por un charco de agua porque el charco no existía, y que ahora, cuando digo estar en lo cierto, mi experiencia visual sí que haya sido causada por una situación mundanal en la que no hay charco alguno. Quien quiera explicar el origen

del concepto de causa debería, seguramente, empezar por aquí o, al menos, tener muy en cuenta este punto. En otras palabras, para tener experiencia del error hace falta que alguna vez alguna de mis experiencias visuales haya sido causada eficientemente por un objeto independiente de la existencia de mi estado psicológico y, precisamente, por el objeto físico que digo estar viendo/percibiendo. Expresado en términos clásicos, lo que estoy diciendo en el fondo es que los datos de los sentidos no son siempre meros signos, manifestaciones o fenómenos de las cosas trascendentes, y que hablar de percepciones inmanentes y trascendentes no tiene mucho sentido para el caso peculiarísimo de las percepciones externas.

En un cierto sentido, el argumento que presento es semejante al que ofrecía Descartes en la tercera Meditación para demostrar la existencia de Dios por los efectos, sólo que en este caso ni siquiera hace falta demostrar esta existencia para demostrar que en un momento determinado yo he conocido un mundo que existe independientemente de mí. Decía Descartes en esa Meditación central:

Y aunque pueda ocurrir que de una idea nazca otra idea, ese proceso no puede ser infinito, sino que hay que llegar finalmente a una idea primera cuya causa sea como un arquetipo en el que esté formal y efectivamente contenida toda la realidad o perfección que en esa idea está sólo de modo objetivo o por representación. (Descartes 1967, 197).

De la misma manera, no puede suceder que, puesto que tengo experiencia del error, mis experiencias visuales hayan sido causadas *siempre* "por mí mismo" al modo del sueño o la alucinación, sino que, al menos alguna vez, he estado *anclado* en la verdad y he reconocido como existente un trozo de mundo tal cual él era. Eso es todo lo que demuestra este argumento y es todo lo que, de hecho, consigue demostrar Descartes en sus famosas Meditaciones. Recuérdese que sólo pudo demostrar que todas

las ideas de las cosas sensibles que se producen en él pasivamente no pueden ser causadas siempre por él mismo, sino que alguna vez han debido ser causadas por algo que contenga formalmente lo que en él está sólo objetivamente. Ni su argumento ni el mío demuestran, sin embargo, que aquí y ahora, en virtud del cúmulo de experiencias y sucesos disponibles para mí en este momento, yo no haya podido salir de la verdad sin darme cuenta de ello, como cuando antes juzgaba que allí delante había un charco y no lo había. En este respecto, la paradoja de las Meditaciones Metafísicas de Descartes es que aunque Dios sea garante de la verdad, sigue siendo posible que aquí y ahora Descartes esté equivocado cuando juzga acerca de la existencia de los objetos particulares en tanto que causas de ciertas experiencias visuales particulares suyas. Por este motivo pudo escribir la primera meditación sobre la duda.

Lo que importa destacar, por lo curioso que resulta, es que, por una parte, la reflexión sobre la experiencia del error y sus condiciones de posibilidad obliga a reconocer que quien realiza y comprende esa peculiar experiencia debe haber estado anclado alguna vez en la verdad y, por otra, que se puede salir en cualquier momento de la verdad sin saberlo. La condición cognitiva subjetiva humana con el mundo de los objetos sensibles es así de curiosa. Por consiguiente, la determinación de estas existencias particulares concierne más bien a la coherencia de las experiencias y efectos a lo largo del tiempo. La idea de que pudiéramos vivir siempre en un sueño constante, o en un estado alucinatorio permanente es absurda, porque entonces no habría posibilidad de discernir el sueño de la realidad y, en este caso peculiarísimo, la falta de criterios es una demostración de que no estamos soñando. De otra manera, que existencia natural es aquello que determina y posibilita el curso coherente de las experiencias sensibles conscientes y la explicación del error.

& 11. Sobre percepción y existencia en la fenomenología.

Por su oportunidad, quiero abrir aquí un breve paréntesis sobre un tema fundamental de la fenomenología husserliana. Mi intención no es otra más que la de hacer unas indicaciones críticas acerca de las relaciones entre percepción y existencia al hilo de unos textos centrales de Husserl que pretendo comentar a la luz de las explicaciones y argumentos previos. Sin embargo, estas indicaciones afectan a una pieza fundamental de toda la teoría fenomenológica, a saber, a la tesis de que el mundo no se me da absolutamente y que, por consiguiente, la tesis del mundo es *siempre* una tesis contingente (Husserl 1950, 108-109). Esta doctrina de la fenomenología trascendental tiene su base en la afirmación de que la contingencia del mundo puede ser afirmada con certeza desde el momento mismo en que se repare en que "toda cosa dada en persona (alles leibhaft gegeben Dingliche) puede no existir" (Ibid.), aunque se nos dé descriptivamente de esta manera peculiar, mientras que las vivencias, cuando son el objeto de actos de reflexión, no pueden no existir. La existencia de estas últimas se me da absolutamente con certeza indubitable. Desde un punto de vista sincrónico estrictamente fenomenológico, tanto en una percepción como en una alucinación el objeto se me da él mismo en persona; lo veo. Según las reflexiones de los párrafos anteriores esta afirmación, a su vez, no querría decir otra cosa más que lo siguiente: los momentos de carácter visual de las vivencias perceptivas y de las vivencias alucinatorias, cuando estas vivencias se aíslan de toda otra vivencia intencional y de su contenido, son muchas veces el origen o disparadero psicológico de un juicio existencial implícito. En la percepción, en el sentido normal de la palabra, la cosa no se me da sólo en persona sino que, además, la pongo como realmente existente, como dice Husserl (1950, 274). Pero esto mismo también puede suceder en una alucinación. Si esto es así, entonces ¿cómo

distinguir una percepción "verdadera" de una alucinación? La respuesta de Husserl, hasta un cierto punto, es igual, o muy semejante, que la que he propuesto en los párrafos previos:

Medito, por lo pronto, como hombre "ingenuo". Veo y cojo la cosa misma en persona. Sin duda que a veces me engaño, y no sólo respecto de las propiedades percibidas, sino también respecto de la existencia misma. Sucumbo a una ilusión o alucinación. La percepción no es entonces una "auténtica" percepción". Pero si lo es, y esto quiere decir que se deja "confirmar" en el curso actual de la experiencia, eventualmente con la ayuda del correcto pensar empírico, entonces existe la cosa percibida realmente y está dada en la percepción realmente ella misma y en persona. (Husserl 1950, 89).

Como también he tratado de argumentar en los párrafos previos, la experiencia a lo largo del tiempo es fundamental para establecer la existencia en general de los objetos sensibles particulares. La importancia que otorga el propio Husserl a la experiencia sensible no se puede dejar de destacar suficientemente. Ella es, por lo pronto, "la fuente última de que se alimenta la tesis general del mundo" (Husserl 1950, 88). Pero es que ella también es la especie de experiencia primaria que fundamenta todas las demás especies de experiencias, incluso debemos suponer que también la--según Husserl mal llamada--"experiencia interna" (Ibid.).

Hasta aquí parece ir todo muy bien. Las cosas, sin embargo, se le complican a Husserl a partir de ahora. Algunos párrafos posteriores a este último sobre el que he llamado la atención afirma Husserl algo que no puede dejar de ser problemático y que, incluso, parece encerrar en buena medida una cierta contradicción con el anterior.

...es inherente, como sabemos, a la esencia del mundo de las cosas que en ninguna percepción, por perfecta que sea, se dé dentro de su dominio un absoluto, y con esto se halla en relación esencial el que toda experiencia, por vasta que sea, deje abierta la posibilidad

de que lo dado, a pesar de la conciencia constante de su estar presente a sí mismo, en su propia persona, no exista. (1950, 108).

Lo que yo afirmo, por el contrario, es que en una percepción "de verdad" tiene que estar dado un "absoluto", una existencia, aunque esta existencia sólo se pueda establecer en el curso mismo de la experiencia posterior de una forma sumamente compleja y sutil. (El hecho de que la cosa percibida se nos dé en escorzos no afecta para nada esta afirmación.) Durante este curso de la experiencia posterior se da "un influjo de las posiciones empíricas posteriores sobre las anteriores" (1950, 108), influjo que consiste enteramente en la posibilidad de que lo que yo juzgaba antes que existía no existiera, como sucedía con los charcos de agua manchegos, y viceversa. Precisamente, en esta posibilidad se encerraría, según Husserl, un criterio especialmente valioso para delimitar la esfera de las existencias absolutamente dadas de las existencias contingentes y dubitables.

En la esfera absoluta no hay espacio para la pugna, la falsa apariencia, el ser de otra manera. Es la esfera de la posición absoluta. (1950, 108).

La consecuencia inevitable de este pensamiento es que la tesis del mundo tiene que ser una tesis contingente, porque toda experiencia sensible, en su conexión con la cuestión de la existencia de los objetos sensibles mundanos, está atada a la posibilidad de la pugna y de la falsa apariencia¹². Sin embargo, esta conclusión es un tanto precipitada a mi

¹² Se puede decir, desde luego, que la existencia del mundo es siempre dubitable según este baremo y considerar que existencia es algo que sólo se puede dar absolutamente en la reflexión sobre los propios estados intencionales. Sin embargo, ¿por qué habría de ser esto así? Ciertamente, se puede dar una definición de existencia como ésta, pero nada impide dar *otra* definición de existencia. Antes ofrecí esta otra de existencia natural: existencia natural es aquello que determina y posibilita el curso coherente de las experiencias sensibles conscientes y la explicación de la experiencia del error. Con este nuevo listón, la existencia natural se da tan absolutamente como

modo de ver, porque del hecho de que a veces, incluso muy frecuentemente, quepa el error a la hora de determinar las existencias de las cosas sensibles particulares, no se sigue necesariamente que *siempre* se pueda afirmar que estas existencias *nunca* se dan absolutamente. Una cosa es que estemos sujetos al error cuando las establecemos y otra bien distinta es que *nunca* podamos establecerlas con seguridad o que podamos no haber estado nunca en contacto epistemológico directo y privilegiado con un trozo del mundo que realmente existe. Desde luego, la existencia indubitable en mí de la experiencia del error y de su sentido hace que la tesis del mundo no pueda ser siempre una tesis contingente como afirma Husserl. La experiencia del error, como experiencia cuyo sentido entiendo perfectamente y que se me da absolutamente hace que, al reflexionar sobre sus condiciones de posibilidad, no tenga otra opción más que la de defender que la tesis del mundo no es en absoluto siempre contingente. Si la tesis del mundo fuera contingente hasta el fin no habría percepciones, ni tendría sentido hablar de la experiencia del error, pero es indubitable que ambas cosas se dan. En especial, la posibilidad de la duda, la posibilidad de decir que el mundo nunca se me da absolutamente, depende esencialmente de la experiencia del error. Pero la posibilidad de la experiencia del error hace que, curiosamente, no se pueda afirmar siempre que el mundo no se me da absolutamente. Alguna vez ha tenido que dárseme así.

la existencia de las vivencias intencionales. Si no se pide más de la existencia natural--y no hay razones para pedir más de ella--no hay motivos para sentirse decepcionados. Sólo cuando se pide otra cosa de la existencia natural, por ejemplo, que se dé absolutamente *en cada ahora del tiempo*, es cuando ésta se nos queda menguada frente a la existencia psíquica y podemos decir que dudamos o que podemos dudar siempre de que se dé. Pero la existencia natural siempre queda asegurada. En cualquier caso, es cierto que la existencia del objeto natural y la existencia de la experiencia visual son, como dice Pears, dos existencias distintas (1976, 29).

La consideración que se acaba de hacer pone también en claro que no cabe concebir prueba alguna sacada de la contemplación empírica del mundo (die Erfahrungsbetrachtung del Welt) y capaz de cerciorarnos con absoluta seguridad de la existencia de éste. El mundo no es dudoso en el sentido de que hubiese motivos racionales que entrasen en cuenta contra la enorme fuerza de las experiencias coherentes, sino en sentido de que es concebible una duda, y lo es porque *jamás* queda excluida la posibilidad de la no existencia, por ser una posibilidad de principio. (Husserl 1950, 109) (El subrayado es mío).

El argumento que he presentado, desde luego, pretende establecer que precisamente de la contemplación empírica del mundo más la experiencia del error que se nos da en el transcurso de esta contemplación, podemos cerciorarnos con absoluta seguridad de la existencia de éste, aunque podamos estar siempre sujetos a múltiples errores particulares. Como decía el propio Husserl en otro contexto--a propósito del escepticismo a que lleva irremediamente el historicismo como doctrina filosófica universal--no se puede dudar de la "verdad de hecho" de que, efectivamente, nos equivocamos o estamos sujetos a error cuando establecemos la existencia de las cosas mundanas. Sin embargo, lo que se trata de establecer en este momento es si lo que Husserl dice se puede justificar con "universalidad de principio" (Husserl 1965, 50-51). Que dentro de ciertos conjuntos de experiencias coherentes dadas en un intervalo de tiempo quepa concebir la duda sobre todas las afirmaciones de existencia que hagamos en ese intervalo temporal acerca de los objetos sensibles particulares--afirmaciones a las que nos ha impelido psicológicamente vivir esas experiencias--, no equivale, ni mucho menos, a pensar que la posibilidad de la no existencia del mundo sea una posibilidad de principio. El curso temporal de las vivencias y la aparición en él, tarde o temprano, de la experiencia del error, dada en su existencia y en su sentido en algún ahora del tiempo,

elimina desde un inicio esa posibilidad de principio. Por muy extraña que parezca esta afirmación, las cosas tienen que ser así.

Si este argumento tiene cierta plausibilidad, entonces el objetivo último de las reflexiones de Husserl, que él creyó haber alcanzado sin el menor asomo de duda, no parece tal fácilmente asequible.

Nuestras consideraciones han llegado con esto a una cima. Hemos logrado los conocimientos de que teníamos menester. En las relaciones esenciales que se nos han abierto están encerradas ya las premisas más importantes para las consecuencias que queremos sacar sobre la posibilidad de principio de desligar el mundo natural entero del dominio de la conciencia, de la esfera del ser de las vivencias; consecuencia de las cuales, como de ello podemos convencernos, se hace finalmente justicia a cierto núcleo de las Meditaciones de Descartes (1950, 109-110).

Mi tesis, una vez más, es que el mundo de la conciencia sensible tiene que estar--¿cómo decir?--*clavado* o *anclado* en el mundo natural de la misma manera en que el taxidermista clava y fija el insecto con un alfiler en el corcho, según la famosa metáfora. La aparición misma en el dominio de la conciencia del sentido y de la experiencia del error--ambas cosas dadas absolutamente--fuerzan esta conclusión. En este sentido, y aquí acabo con este parágrafo a modo de comentario, la reflexión sobre la percepción y la existencia, sobre la experiencia sensible y la experiencia del error, hacen que una de las tesis máximas de la fenomenología trascendental no pueda ser verdadera con los argumentos que Husserl presenta: que la tesis del mundo es siempre contingente y que, por ello, el mundo de la conciencia

se puede desanclar del mundo natural. Sencillamente no parece que eso sea así para el caso de la conciencia perceptiva.

& 12. Los datos de los sentidos y el conocimiento del mundo externo.

Si de algo han servido todas estas reflexiones ha sido para establecer que la posibilidad de dudar siempre de la existencia de todos los objetos particulares que se me dan por los sentidos no es una posibilidad de principio, si es que el problema del conocimiento al estilo cartesiano quiere empezar siquiera a plantearse. Por consiguiente, en las argumentaciones que siguen me serviré con generosidad de la afirmación de que podemos contar con la existencia general de los objetos sensibles mundanos y con la de sus propiedades.

Como complemento al problema de la percepción y del error, y por lo que respecta al caso de los informes verbales de los que se hablaba para dilucidar el concepto de conocimiento por familiaridad, podría suceder desde luego que lo que dice mi amigo acerca del monstruo que vive debajo de la superficie helada del lago Mendota fuera falso. Podría ser un mentiroso o estar mal informado por algún motivo. Tanto en un caso como en otro hay experiencia del error; en este caso experiencia de *su* error.

Sin embargo, y para volver al problema de la percepción, dado que es posible que la manzana no esté sobre la mesa—incluso que la mesa misma tampoco esté ahí—, no podemos decir sin ninguna clase de reserva crítica cuando nos mantenemos aislados en la vivencia de la experiencia sensible que es constituyente imprescindible de las percepciones y las alucinaciones, que siempre tenemos conocimiento por familiaridad de la manzana misma cuando decimos estar viéndola, al menos según los baremos estrictos que establece Russell para esta clase de conocimiento. Como la

reserva crítica en estas condiciones es procedente y como, en cualquier caso, seguimos teniendo ciertas experiencias sensibles, tanto si estamos en un estado perceptivo como en uno alucinatorio, parece plausible reconocer que debemos tener familiaridad con alguna otra clase de entidad diferente, que no es la manzana misma, siempre que tengamos una experiencia sensible. Russell llama a esta otra entidad "datos de los sentidos" (sense data), categoría a la que pertenecerían, por ejemplo, los sonidos que oímos, los colores que vemos organizados espacialmente, la dureza de los objetos que sentimos mediante el tacto, etc. Esta peculiar clase de objetos, los datos de los sentidos, a la que Brentano llamaba "fenómenos físicos", se destaca porque somos directamente conscientes de ellos en sus correspondientes especies de sensaciones, y porque su conocimiento es infalible y completo. Cuando yo digo que veo una manzana es que la veo, aunque esté en un estado alucinatorio. Por el contrario, nuestro conocimiento de la existencia de los objetos del mundo exterior, cuando se acota en un intervalo de tiempo breve, no parece ser siempre ni infalible ni necesariamente completo. Podemos salir de la verdad sin darnos cuenta de ello.

Hay que señalar, por otra parte, que los datos de los sentidos no son los únicos candidatos que hay para formar parte del círculo restringido de las cosas respecto de las cuales tenemos conocimiento por familiaridad. No hay simplemente un campo de datos de los sentidos más o menos entrelazado y unitario--esté o no causado por un objeto realmente existente--del cual tenemos conciencia. Los colores que vemos, los sonidos que oímos, etc. son, además, "interpretados" o "aprehendidos" de determinadas maneras. Por ejemplo, éstos que ahora experimento son aprehendidos como una manzana encima de la mesa. Y de esta "aprehensión" o "interpretación" de los datos de los sentidos también tenemos necesariamente conocimiento por familiaridad. Haya o no haya manzana "allí fuera en el mundo extralingüístico", yo digo que veo una manzana, aunque todo sea un puro

espejismo. De que yo veo una manzana tengo conocimiento por familiaridad. Nadie me puede apear de este burro. Por lo tanto, en todo fenómeno visual, bien sea parte de una percepción o de una alucinación, es necesario distinguir, cuando menos, estos dos elementos imprescindibles: los datos de los sentidos y la "interpretación" o "aprehensión" que hago de esos datos, a los que habría que añadir el elemento indexical del que se habló anteriormente. De todos estos elementos tenemos conocimiento por familiaridad.

¿Qué sucede, sin embargo, con el conocimiento de las manzanas en tanto que objetos físicos existentes, y no meramente con el conocimiento de ciertos datos de los sentidos aprehendidos como una manzana que está allí sobre la mesa? Como se desprende fácilmente de los comentarios previos, es importante distinguir entre percepción y experiencia visual simplemente. En una percepción, por ejemplo, tenemos una experiencia de tipo o cualidad visual, y el objeto que percibimos existe y es tal y como decimos que es, porque es el objeto que ha causado la aparición en el río de la conciencia de la experiencia visual que es parte inseparable de esa percepción. Por supuesto, puede suceder que tengamos una experiencia visual, pero no objeto externo que, de hecho, se ajuste a nuestra experiencia visual. Eso es una alucinación, como bien se sabe.

Ahora bien, hay una relación interesante entre los informes verbales de los sujetos y sus experiencias visuales--percepciones o alucinaciones. Esta relación se puede describir sucintamente de la siguiente manera. Pensemos en uno de los ejemplos anteriores. Si soy un hablante competente, tengo familiaridad con el significado de las palabras y enunciados que pronuncia mi amigo y con el estado de cosas al que él intenta referirse cuando dice que "hay un monstruo que vive debajo de la superficie helada del lago Mendota". Dado que es perfectamente posible concebir que mi amigo esté mintiendo o equivocado, no puedo estar familiarizado con un estado de

cosas actual, ahí en el mundo, del que también yo informaría diciendo que "hay un monstruo que vive debajo de la superficie helada del lago Mendota" cuando meramente entiendo sus palabras y no tengo ninguna percepción (verdadera) de ese monstruo.

De todo lo anteriormente dicho se desprende que debemos tener familiaridad con nuestros datos de los sentidos, con su "aprehensión" o "interpretación"--cuando se produce--, con el significado de las palabras y enunciados, incluidos los deícticos, si somos hablantes competentes, y con ciertos estados de cosas posibles que nos representamos inevitablemente cuando entendemos el significado de las palabras y oraciones. Probablemente, el problema del conocimiento no es el de si tenemos un conocimiento verdadero de estas cuatro clases de entidades. El problema del conocimiento, expresado en toda su amplitud y profundidad, es que queremos conocimiento verdadero, objetivamente válido, entre otras cosas, *de los objetos que hay en el mundo*. Queremos saber cuándo estamos plenamente justificados para afirmar que tenemos percepciones de objetos externos y cuándo son verdaderos nuestros juicios de existencia acerca de estos objetos. No sólo queremos afirmar que tenemos conocimiento infalible y verdadero acerca de un conjunto de datos de los sentidos del que cada uno de nosotros informaría diciendo, por ejemplo, que "yo veo una manzana", sino que también queremos saber cómo es posible que tengamos conocimiento verdadero de la manzana en tanto que objeto externo independiente de nuestra conciencia y del río temporal incesante de nuestras vivencias.

Sin embargo, hay una complicación añadida. Imagínese que mi amigo tiene una alucinación y que, por su causa, cree firmemente que está viajando en un submarino bajo la superficie helada del lago Mendota y que está, de hecho, percibiendo un monstruo en ese momento. Yo estoy junto a él en su casa cuando me informa sinceramente de su experiencia visual

de esta manera: "veo un monstruo que vive debajo de la superficie helada del lago Mendota". Imagínese que yo no padezco ninguna alucinación y que sé que no hay tal monstruo. En este caso, mi amigo, desde luego, está familiarizado con un conjunto de datos de los sentidos, con su "interpretación" y con el significado de las palabras que él usa en el informe verbal de dicha experiencia. En estas circunstancias, yo estaría dispuesto a conceder que el estado de cosas al que él está intentando referirse cuando usa aquellas palabras tiene para él el sentido de *ser un monstruo real* que vive debajo de ese lago. En la intención de sus palabras hay implícito un juicio de existencia acerca de la realidad de cierto monstruo. Por "detrás" de su enunciado hay operando un juicio existencial, y este juicio existencial determina la intención y el uso de sus expresiones al igual que, muy probablemente, su comportamiento. Como he argumentado, su experiencia visual funda o causa un juicio existencial de esta manera. Con todo, debido a mi experiencia de *su* error, estoy en condiciones de decir que el estado de cosas al que él se refiere realmente debería identificarse más bien con sus datos de los sentidos interpretados de determinada manera y no con monstruo real alguno. Para mí él debería haber dicho algo así como "*me parece* percibir un monstruo que vive debajo de la superficie helada del lago Mendota", y no que "*hay* un monstruo que vive debajo de la superficie helada del lago Mendota". Aplíquese el mismo razonamiento para el caso de don Quijote y Sancho.

Para explicar técnicamente esta situación, es necesario introducir aquí un poco de terminología nueva. Siguiendo a Husserl, llamaré "representaciones impropias o simbólicas" a aquellas representaciones en las que significamos directamente un objeto, pero en las que, como diría Russell, no estamos familiarizados con él (Husserl 1970, 193). En una representación simbólica estamos familiarizados con un objeto, pero este objeto no es el que mencionamos explícitamente con esa representación. En

este sentido, los datos de los sentidos y, quizá, los significados son, al menos en algunas ocasiones, representaciones simbólicas de objetos físicos y estados de cosas. Por ejemplo, los datos de los sentidos son representaciones simbólicas cuando tenemos familiaridad con estos datos de los sentidos pero tenemos la intención o propensión natural, determinada por una "irresistible ley de nuestra naturaleza" como decía Portenfield, de referirnos mediante ellos a un objeto que está en el mundo. Dado que estamos familiarizados con los datos de los sentidos, pero que tenemos la intención, prácticamente constante, de referirnos a un objeto que es distinto, en su modo ontológico, de los datos de los sentidos, los datos de los sentidos son representaciones simbólicas de este objeto. Sin embargo, los datos de los sentidos no son representaciones simbólicas cuando no tenemos la intención de referirnos a "través de ellos" a un estado de cosas actual en el mundo. Si yo digo "me parece percibir una manzana" y esta oración es la expresión de una descripción verdadera de un conjunto de datos de los sentidos interpretados de determinada manera y que estoy experimentando en este momento, entonces estos datos de los sentidos no son representaciones simbólicas, porque no tengo la intención de juzgar acerca de la existencia de una manzana en el mundo exterior. Mi juicio lo es acerca de mis datos de los sentidos así aprehendidos, y no acerca de un objeto externo. En este caso, los datos de los sentidos son, en la reflexión, representaciones propias de sí mismos.

Merece la pena señalar que, sin embargo, casi nunca estamos *explícitamente* familiarizados con los datos de los sentidos cuando decimos tener una percepción; que el modo por familiaridad del conocimiento, curiosamente, no se nos da nunca inmediatamente, a menos que reflexionemos acerca de nuestras experiencias visuales y pensemos sobre las de los demás sobre la base de la experiencia del error. La noción de dato de los sentidos sólo se origina en la reflexión sobre el error: sobre el error

propio y el ajeno. No hablamos nunca acerca de los datos de los sentidos, *qua* datos de los sentidos, en nuestros enunciados perceptivos (Castañeda 1977, 292). Si no hubiera experiencia del error jamás habiéramos llegado a tener idea semejante. Después de realizar esta experiencia, se puede realizar la reflexión todas las veces que se quiera y pensar que uno está hablando acerca de los datos de los sentidos y no acerca de objetos en el mundo. Sin embargo, sólo cuando se ha percibido alguna vez de verdad un objeto, es decir, sólo cuando ha habido una correspondencia contrafactual--por consiguiente causal--entre la existencia y propiedades de este objeto, y la existencia y propiedades de nuestra experiencia visual subjetiva, de tal manera que dicha experiencia no hubiera tenido nunca ninguna de sus propiedades si es que el objeto no hubiera existido así y así--, es cuando es posible tener experiencia del error y cuando se puede encontrar en la reflexión los "datos de los sentidos" como entidades de las que tenemos conocimiento por familiaridad¹. En la vida de todos los días, los datos de los sentidos no son en absoluto los objetos sobre los que queremos juzgar y de los que hablamos. Russell distingue en este sentido entre los objetos físicos, que supuestamente son la causa de nuestros datos de los sentidos, y los objetos directamente sensibles o datos de los sentidos.

De acuerdo con estas indicaciones, en nuestro uso normal de los enunciados, de las descripciones definidas y de los nombres propios, tratamos de referirnos a los objetos físicos del mundo y no a los datos de los sentidos. En la vida cotidiana consciente, los datos de los sentidos, interpretados "automáticamente" así y así, tienen una estructura intencional o de "flecha", por así decir. Los datos de los sentidos tienen esta estructura

¹ Es por este motivo por el que no se puede empezar sin más una teoría del conocimiento por los datos de los sentidos de un modo atomista como hizo, por ejemplo, Locke.

de "flecha" porque funcionan constantemente como representaciones simbólicas con las que no estamos explícitamente familiarizados.

El significado de las palabras y enunciados seguramente tiene la misma clase de estructura. En la vida de todos los días, mediante el uso que hacemos de los informes verbales que expresan el contenido de nuestras experiencias visuales tratamos de referirnos a objetos físicos existentes, pero no a los datos de los sentidos que constituyen esencialmente nuestras experiencias visuales. Sin embargo, cuando no reflexionamos sobre nuestros datos de los sentidos, ellos no son específicamente el objeto de nuestras percepciones internas y de nuestros informes verbales. La estructura de "flecha" de los datos de los sentidos, *qua* datos de los sentidos, es "absolutamente transparente", por decirlo de algún modo gráfico e intuitivo, y nos lleva a pensar de una manera "natural" que estamos familiarizados con el mundo exterior.

Como he señalado anteriormente, puede no haber diferencia entre la experiencia visual como parte integrante de una alucinación y la experiencia visual que es parte integrante de una percepción, en cuanto a la manera en que se nos dan los datos de los sentidos que estamos experimentando y a cómo los utilizamos para inferir juicios existenciales. En ambos casos, tenemos la intención de mencionar el mismo objeto físico y no los mismos datos de los sentidos. Pero como dije, dado que tenemos experiencia del error y de su sentido, podemos llegar a dudar de si nuestras menciones puntuales de objetos en la vida cotidiana se ajustan, de hecho, a la realidad, a lo que hay. Pero aunque podemos tener estas dudas razonables, podemos explícitamente reflexionar sobre nuestras experiencias visuales y centrar toda nuestra atención sobre los datos sensibles, de tal forma que estos datos sensibles resulten ser ahora los objetos propios de nuestra percepción interna. Cuando los datos sensibles son los objetos de un acto de reflexión, decimos que estamos familiarizados explícitamente con estos

datos sensibles, y podemos hacer informes lingüísticos verdaderos acerca de ellos. En este caso, los datos sensibles no son representaciones simbólicas nunca más. El concepto de dato sensible es, pues, una "idea de reflexión" y no de "sensación" que se conoce después de haber explicado la experiencia del error en sus condiciones de posibilidad esenciales. Cuando los datos sensibles son explícitamente los objetos de un acto de reflexión, ellos pierden temporalmente su estructura de flecha transparente y el mundo parece quedar en suspenso para nosotros en su realidad objetiva. La reflexión, al dirigir la atención exclusivamente sobre los datos de los sentidos, ha cortado de pronto el "disparadero" de todos nuestros juicios existenciales, y que no es otro que el dejar operar a nuestra conciencia en toda su riqueza de operaciones judicativas. Cuando informo acerca de la estructura de mis datos de sensación y digo "me parece ver una manzana", no estoy diciendo que "hay una manzana ahí" en un "ahí" localizable intersubjetivamente. En este caso, cualquier sujeto que reflexione sobre sus datos de los sentidos tiene la oportunidad de dejar de estar comprometido con lo que hay en el mundo. Pero tan pronto como pierda la actitud reflexiva acerca de los datos sensibles y deje de estar cohibido judicativamente acerca de lo que hay, ellos ganan de nuevo su estructura de flecha "sin pensarlo" y recuperan inmediatamente el papel de ser representaciones simbólicas "transparentes" no explícitamente representadas como tales, de los objetos físicos de la vida natural cotidiana.

De estas reflexiones parece extraerse la conclusión siguiente. Hasta cierto punto, esto es, alejados de la actitud crítica de reserva que impone la reflexión que dirige toda la atención sobre los datos de los sentidos, podríamos decir que estamos también familiarizados con el objeto físico que nuestros datos de los sentidos mencionan en virtud de su estructura de flecha cuando no reflexionamos sobre ellos y *estamos teniendo realmente una percepción*. Hablaré de "familiaridad impropia" cada vez que quiera

mencionar el modo en que se nos dan los objetos externos en virtud de la estructura de flecha de nuestros datos de los sentidos en la vida cotidiana. Por supuesto, es posible que el objeto que explícitamente mentamos con nuestras representaciones simbólicas no exista. Sin embargo, merece la pena señalar de nuevo que esta circunstancia no es necesariamente siempre el caso. Es más, que no puede ser siempre el caso. Recuérdese toda la argumentación anterior acerca de los charcos de agua. El objeto que se menciona en una representación simbólica--al vivir ciertos datos de los sentidos--puede ser tal y como decimos que es, porque es el objeto que ha causado la existencia de esa representación.

El extraño concepto de familiaridad impropia es, como puede apreciarse muy bien, una consecuencia de las características de nuestra relación epistemológica peculiar con el mundo de los objetos físicos. La familiaridad impropia es simplemente una forma abreviada de decir que no es una posibilidad de principio dudar siempre de la existencia de los objetos sensibles mundanos particulares, aunque siempre cabe la posibilidad de "salir de la verdad" sin saberlo.

Evidentemente, a Russell le está vedada enteramente la posibilidad de contar con una noción semejante a la de familiaridad impropia. Por ese motivo, su noción de familiaridad y la mía diferirán en un aspecto esencial a partir de este momento. El no puede disponer libremente de la idea de familiaridad impropia, porque no piensa que las representaciones simbólicas--los datos de los sentidos o los significados--tengan una estructura intencional, como la tienen para Frege o para Husserl. Para Russell sólo hay una relación directa de referencia o de *familiaridad propia*. No hay, en adición a esto, una relación mediada con algo con lo que uno pudiera tener un contacto semántico o epistémico "indirecto". En particular, la relación del significado de una descripción con su denotación no es para Russell la relación intencional de una representación impropia o simbólica

con lo que está representado; no es una relación inherentemente intencional o representacional².

Resumiendo, y según la teoría y las distinciones que propongo, tenemos conocimiento por familiaridad de las siguientes entidades:

(i) Estamos explícitamente familiarizados con los datos de los sentidos en actos de reflexión e implícitamente cuando funcionan como representaciones simbólicas transparentes.

(ii) Estamos siempre familiarizados con las "interpretaciones" o "aprehensiones" de estos datos de los sentidos.

(iii) Estamos familiarizados con los significados de las palabras y oraciones cuando somos hablantes competentes.

(iv) Estamos familiarizados con estados de cosas posibles que nos representamos cuando entendemos el significado de las palabras y oraciones.

(v) Tenemos "familiaridad impropia" con el objeto externo al que aludimos en virtud de la estructura de flecha transparente de nuestros datos sensibles, y en virtud de la estructura de flecha del significado de las palabras y enunciados que expresan el contenido de esos datos sensibles, cuando dicho objeto ha sido precisamente el causante de la experiencia sensible de una percepción.

Con estas herramientas terminológicas a la mano, estamos en buena posición para analizar ahora algunas afirmaciones que Russell hace acerca del conocimiento por familiaridad y el conocimiento por descripción. Russell dice que "en la percepción tenemos familiaridad con el objeto de percepción, y en el pensamiento tenemos familiaridad con un objeto de un carácter más abstracto. Pero no tenemos *necesariamente* familiaridad con el objeto que es denotado por frases compuestas de palabras con cuyo significado estamos familiarizados" (1984, 41-42; el subrayado es mío). Ahora

² Debo esta aclaración a la profesora Leora Weitzman de la Universidad de Wisconsin-Madison.

es posible parafrasear esta tesis de esta forma más precisa: tenemos explícitamente familiaridad con los datos sensibles en actos de reflexión que son "disparados" por la experiencia del error, y en el pensamiento estamos familiarizados con el significado de las palabras y enunciados, y con los estados de cosas posibles que nos representamos cuando entendemos el significado de las palabras y oraciones; pero tenemos familiaridad impropia, cuando se da una percepción, con el supuesto objeto externo que tengo la intención de mencionar en virtud de la estructura de flecha de mis datos sensibles y de la estructura de flecha de las palabras y oraciones con cuyos significados estoy familiarizado.

A partir de este texto de Russell, podemos extraer la conclusión de que no tenemos *necesariamente* familiaridad con el objeto denotado por frases compuestas de palabras con cuyos significados estamos familiarizados, porque algunas veces tenemos meramente familiaridad con el significado, por ejemplo, de una descripción definida que lo denota y no con el objeto mismo. Así, el significado de la descripción "el monstruo que vive debajo de la superficie helada del lago Mendota" es un objeto con el que estamos familiarizados, aunque no lo estemos obligatoriamente con el objeto denotado por esa frase. Ya que algunas veces usamos una descripción definida con la intención de seleccionar un único objeto físico de entre otros muchos, estaremos familiarizados con los objetos físicos si es que tenemos conocimiento verdadero de estos objetos, es decir, si es que tenemos una percepción de este objeto físico.

Si esta interpretación no es completamente errónea, esta otra tesis de Russell no puede ser del todo cierta:

... parecerá que entre el objeto con el que estamos familiarizados no están incluidos los objetos físicos (como opuestos a los datos de los sentidos), ni las mentes de otras personas. Estas cosas son conocidas por nosotros por lo que yo llamo "conocimiento por descripción". (1956, 214).

& 13. Percepción y descripciones definidas.

Es esencial distinguir entre una descripción definida como "la actual reina de Inglaterra" y el objeto singular denotado por esa descripción. Llamaré a este objeto singular, como es canónico hacer, la "denotación" o el "referente" de una descripción definida. Es claro que una expresión, *por su forma lingüística*, puede ser una expresión denotante y, sin embargo, no tener ninguna denotación. Si alguien dijo "el actual rey de Inglaterra" en 1992 d.C. en un "discurso de actualidad" (Donnellan 1974, 5), su descripción fue una frase denotante, pero que no tenía ninguna denotación. Así era las cosas en el mundo en ese momento. Decir que una frase es denotante pero que no tiene denotación es decir lo mismo que el hablante, en el uso de esa frase, tiene la intención de referirse a un objeto único existente, pero que no consiguió referirse a él.

Es normal suponer que una descripción definida es una expresión denotante en virtud del significado de las palabras que la componen. Si somos hablantes competentes, tenemos familiaridad con estos significados. En la anterior expresión denotativa sé qué significa "actual", "rey" e "Inglaterra" y su unión sintácticamente bien construida "el actual rey de Inglaterra". Al estar familiarizados con estos significados, tenemos alguna clase de noticia de los posibles objetos que se ajustarían o cumplirían aquella descripción en determinadas circunstancias, aunque no estemos necesariamente familiarizados con ellos. La posibilidad de estar familiarizados con el significado de las palabras y enunciados, en resumidas cuentas, de hablar un lenguaje, y, sin embargo, de no estar familiarizados, ni propia ni impropriamente, con sus denotaciones, es un signo de que no podemos identificar simplemente el significado de una descripción con su denotación.

Dado que estamos familiarizados con el significado de las palabras cuando somos hablantes competentes, Russell está equivocado, así me parece,

cuando piensa que "las frases que denotan nunca tienen un significado en sí mismas, pero que toda proposición en cuya expresión verbal ellas ocurren tiene un significado" (1984, 43). Mi problema es que no puedo entender del todo por qué la frase denotante "el actual rey de Inglaterra" no tiene significado en sí misma, pero que la proposición "el actual rey de Inglaterra es alto" sí que tiene significado, aunque sea falsa. Strawson, por ejemplo, tenía el mismo problema (1985)³.

Por estar familiarizado con el significado de la descripción "el actual rey de Inglaterra", parece más o menos claro que cuando usé estas palabras en un discurso de actualidad en 1992 d.C., tenía la intención de denotar un objeto definido. Dado, por otra parte, que no había ningún rey de Inglaterra en 1992, la frase "el actual rey de Inglaterra" falló a la hora de denotar un objeto cuando la pronuncié en 1992. Así, pues, la descripción definida "el actual rey de Inglaterra", que Russell ofreció como ejemplo de una frase que denota y que denota además un objeto definido, no denota, sin embargo, ningún objeto en 1992.

De acuerdo con estas indicaciones introductorias, necesitamos distinguir, al menos, las siguientes cosas, por lo demás, de sobra conocidas.

³ La pregunta es: ¿por qué Russell dice que la descripción "el actual rey de Inglaterra" no tiene significado por sí misma? Es claro que Russell piensa que las partes que componen lingüísticamente esta expresión tienen significado y que contribuyen a los significados de las oraciones que las contienen. Sigo pensando que Russell hubiera tenido que decir que estas descripciones sí que tienen significado, aunque no denotación. En otro sentido, sin embargo, Russell podría intentar buscar una justificación para su tesis. En lo que realmente quiere insistir Russell es en que el significado con el que la frase denotante contribuye al significado de la oración en la que aparece no puede ser *monolítico*. Su contribución semántica es, más bien, compleja porque añade un cuantificador existencial y un predicado al significado de la oración ("hay uno y solamente un objeto que es el actual rey de Inglaterra y ese objeto es alto"). Creo que Russell tiene miedo de decir que la expresión denotante tiene significado por sí misma, porque de esta manera sus lectores podrían malinterpretarle y pensar que tendría un significado monolítico simple. Este significado, según Russell, tendría que ser o un sentido fregeano, en el que Russell no cree, o el objeto denotado, en cuyo caso cuando nada se ajusta a la descripción dicho objeto tendría que ser un objeto no existente *à la* Meinong. Pero Russell tampoco cree en esta clase de entidades.

(i) La descripción definida "la actual reina de Inglaterra" en tanto que expresión.

(ii) El significado de las palabras "la actual reina de Inglaterra". La descripción "la actual reina de Inglaterra" es una frase denotante porque tiene significado.

(iii) La persona real, esto es, la denotación o el referente: la actual reina de Inglaterra misma.

(iv) El tiempo o contexto en que se pronuncia esta frase denotante.

(v) La intención de denotar a una persona o a un objeto singular cada vez que usamos una descripción definida "en un discurso de actualidad".

La descripción definida "la actual reina de Inglaterra" tendrá denotación dependiendo del tiempo y, en general, del contexto en que sea pronunciada. De acuerdo con estas distinciones básicas, tengo meramente conocimiento por descripción de la actual reina de Inglaterra cuando entiendo el significado de las palabras "la actual reina de Inglaterra" sin por ello tener un verdadero conocimiento por familiaridad de la actual reina de Inglaterra, es decir, sin tener una percepción (verdadera) de esa mujer. Pero si beso la mano de la actual reina de Inglaterra al tiempo que el jefe de protocolo de la casa real inglesa dice "la reina de Inglaterra" mientras me está presentando a ella, y es verdad lo que él está diciendo, y yo tengo una percepción verdadera de la reina, etcétera, entonces tengo realmente conocimiento por familiaridad y conocimiento por descripción de la actual reina de Inglaterra.

Si estas ideas no están completamente fuera de lugar, entonces el siguiente texto de Russell es, cuanto menos, algo desorientador:

por ejemplo, sabemos que el centro de la masa del sistema solar en un tiempo definido es un punto definido, y podemos afirmar cierto número de proposiciones acerca de él; pero no tenemos *familiaridad*

con este punto, que para nosotros es sólo conocido por descripción. La distinción entre *familiaridad* y *conocimiento sobre* es la distinción entre las cosas de las que tenemos representación, y las cosas que sólo alcanzamos por medio de frases que denotan. (1984, 41).

Cuando Russell dice que no tenemos familiaridad inmediata con este punto, tiene que querer decir algo así como que no tenemos una percepción de este punto. Esta afirmación es algo que, desde luego, acepto. En este momento, sólo contamos con una teoría física que nos permite decir que, probable e idealmente, hay un único punto que es el centro de la masa del sistema solar en cada instante de tiempo. Pero el uso por parte de Russell de la palabra "representación" es muy equívoco en este texto, porque nosotros, ciertamente, tenemos una representación de este punto, aunque sea una representación mediada inferencialmente. Como he señalado anteriormente, tenemos una cierta representación de este punto porque entendemos el significado de ciertas palabras y enunciados, y porque sabemos qué es lo que tendría que suceder para que esta frase tuviera una denotación dado, por supuesto, nuestro conocimiento de cierta teoría física. Quizás Russell debería haber dicho más precavidamente que la distinción entre conocimiento por familiaridad y conocimiento sobre, es la distinción que hay entre las cosas con las que estamos familiarizados y los objetos físicos externos de los que no tenemos siempre *necesariamente* una percepción, pero a los que podemos referirnos cuando usamos una descripción definida. En cualquier caso, no parece que la clase de conocimiento por descripción que Russell defiende que tenemos del centro de masa del sistema solar y de los objetos físicos sea de la misma "calidad". Pienso que esta diferencia en su calidad es mejor captada por la teoría que presento que por la de Russell. De algún objeto físico debemos tener necesariamente alguna vez una percepción. Del centro de masa del sistema solar, no necesariamente.

Como sugerí más arriba, Russell sólo puede decir que tenemos conocimiento por familiaridad de los datos de los sentidos y de sus interpretaciones. Sin embargo, he estado argumentando que podríamos llegar a tener conocimiento por familiaridad y conocimiento por descripción del mismo objeto físico en ciertas circunstancias epistémicamente privilegiadas que, como argumenté más arriba, han de darse necesariamente alguna vez. Recuérdense el caso de la presentación formal de la reina de Inglaterra por parte del jefe de protocolo y, sobre todo, el caso de los charcos de agua manchegos. Respecto del último ejemplo de Russell, hay que comentar que si tuviéramos una percepción verdadera del centro de masa del sistema solar-- caso de que pudiéramos percibir un simple punto--, tendríamos conocimiento por familiaridad y conocimiento por descripción de semejante punto. Tenemos conocimiento por descripción de ese punto porque la descripción definida "el centro de la masa del sistema solar" es una representación simbólica. Y es una representación simbólica porque estoy familiarizado con el significado de las palabras que componen esa descripción sin que la presencia efectiva de ese objeto haya sido el origen del uso de semejante descripción, aun cuando mi intención esté dirigida hacia el objeto "físico" que es el centro de la masa del sistema solar. Sólo podemos mencionar de momento objetos como el centro de la masa del sistema solar usando descripciones definidas, porque no tenemos una percepción de este punto.

A este respecto, Russell dice algunas otras cosas que son nuevamente algo desorientadoras. Por ejemplo, hace afirmaciones como ésta:

Lo que deseo discutir es la naturaleza de nuestro conocimiento concerniente a los objetos en los casos en que *sabemos que hay* un objeto que responde a una descripción definida, aunque no estemos *familiarizados* con ningún objeto semejante" (1956, 214. Los énfasis son míos).

Si realmente estamos en posesión de conocimiento, esto es, si sabemos que es verdad *que hay* un objeto que se ajusta a una descripción definida, *necesitamos haber estado previamente familiarizados con semejante objeto* o, al menos, con una teoría de la que pueda deducirse con certeza la existencia del objeto al que se refiere la descripción definida. Por supuesto, bien podría suceder que no estemos familiarizados con este objeto ahora mismo, porque el objeto no nos esté presente en el modo de la percepción. Pero si sabemos que es *verdad que hay* un objeto que se ajusta a una descripción definida, necesitamos haber estado o bien familiarizados con este objeto en el pasado, o tener razones bien fundadas para hacer esa afirmación. Estas razones bien fundadas podrían ser algunas razones basadas en una teoría física bien justificada.

& 14. El rompecabezas de Russell.

Entre la teoría de Meinong, tal y como la formula Russell, que dice que toda frase denotante y que ocurra en una proposición es una frase que está por un objeto, y la teoría de Russell que dice que las proposiciones en las que aparezcan frases denotantes se deberían escribir nuevamente de una manera en la que no aparezca ya más la frase que denota, hay, pienso, un punto intermedio razonable que hay que buscar.

Russell dice que "debemos o bien proporcionar una denotación en los casos en los que está ausente a primera vista, o debemos abandonar el punto de vista de que la denotación es lo que está en juego en las proposiciones que contienen frases que denotan" (1984, 47). Creo que este dilema es falso y que el modo fregeano de resolverlo mediante la distinción entre sentido y referencia es el comienzo de una solución completa del problema. Sin embargo, Russell sostiene que la solución de Frege es "artificial" porque "no contiene un análisis exacto del asunto", aunque

parece que no conduce a errores lógicos de importancia (1984, 47). Para Frege, y para mí, la frase denotante "el actual rey de Francia" tiene significado--ése es precisamente el motivo por el que la frase es una frase que denota--, pero no tiene denotación. Cuando usamos este enunciado, puede que tengamos la intención de denotar un único objeto. En ciertas circunstancias y usando otras fuentes de conocimiento aparte de lo que significa la frase que denota, podemos llegar a saber que no hay actualmente ningún rey de Francia y que, como resultado, nuestra frase tiene significado, aunque no denotación.

Pero, ¿por qué dice Russell que la idea de Frege es artificial y que no proporciona una descripción precisa del problema? Básicamente, porque la propuesta de Frege no es tan buena como la de Russell para resolver algunos rompecabezas lógicos que pueden ser considerados, en este contexto, como experimentos cruciales en cuestiones semánticas. Centraré mi atención en el primero que Russell propone en su célebre artículo "Sobre la denotación" e intentaré ofrecer una solución alternativa a este rompecabezas que haga justicia a Frege, a la vez que a las teorías más recientes de la referencia directa como las de Donnellan y Kripke.

El rompecabezas es expuesto cristalinamente por Russell de la siguiente manera:

Si *a* es idéntico con *b*, lo que sea verdad de uno es verdad del otro, y cualquiera puede ser sustituido por el otro en cualquier proposición sin alterar la verdad o la falsedad de esa proposición. Ahora bien, Jorge IV deseaba saber si Scott fue el autor de *Waverley*; y, de hecho, Scott *fue* el autor de *Waverley*. De aquí que podamos sustituir *Scott* por *el autor de "Waverley"* y, por lo tanto, probar que Jorge IV deseaba saber si Scott era Scott. Sin embargo, un interés por la ley de la identidad difícilmente se puede atribuir al primer gentilhomme de Europa. (1984, 47-48).

Las siguientes secciones están dedicadas a clarificar las complejas relaciones que se dan entre los nombres propios y las descripciones definidas, para sugerir de este modo una solución posible al rompecabezas de Russell.

& 15. Nombres propios y descripciones definidas.

Antes de proponer esta posible solución, sugeriré los siguientes puntos en relación a la teoría de los nombres propios y las descripciones definidas:

(i) Los nombres propios no son ni descripciones disfrazadas, ni son sinónimos con cierto conjunto de descripciones definidas, ni tienen significado, pero están íntimamente asociados con descripciones significativas, definidas e indefinidas.

(ii) Las descripciones que asociamos con los nombres propios tienen significado.

(iii) Los nombres propios no se refieren de por sí a ningún objeto en exclusiva. Son referencialmente inertes. Somos nosotros los que tenemos la intención de referirnos a, o de hablar de, un objeto único cuando *usamos* un nombre propio. Muchas personas tienen el mismo nombre propio, pero nosotros tenemos normalmente la intención de referirnos bien a una bien a otra cuando usamos ese nombre y no a todas ellas.

(iv) Los nombres propios son utilizados por sus usuarios como abreviaturas útiles de un conjunto de descripciones—definidas e indefinidas—que ellos asocian, o estarían dispuestos a asociar en ciertas circunstancias, con dichos nombres. En este sentido se comportan de la misma manera que la etiqueta de un archivo en tanto que ésta es también una abreviatura útil del conjunto de documentos que guardamos en él. (El profesor Michael Byrd me sugirió esta analogía que, como se verá más adelante, es tan

operativa y esclarecedora⁴). El uso de los nombres propios por parte de los hablantes responde, por así decir, a principios de economía y orden cognitivos.

(v) Las descripciones, y, sobre todo, las definidas--los documentos en el archivo--, son criterios que el hablante puede utilizar para identificar la denotación o el referente de un nombre cuando usa este nombre con la intención de referirse a un único objeto. Con esas descripciones cree poder seleccionar ese objeto; precisamente aquel al que quiere referirse mediante el uso del nombre.

(vi) Que *realmente* esas descripciones sirvan como criterios para identificar el objeto al que el hablante quiere referirse depende de si este objeto satisface esas descripciones o, como dice Searle en una versión más relajada del asunto, el número "suficiente" de ellas que sea relevante para seleccionar la referencia o denotación. Esto, a su vez, tal y como se verá inmediatamente, depende esencialmente de la "cualidad" y el origen que las descripciones tengan para cada hablante particular.

Como requisito previo antes de comenzar a aplicar y a desarrollar más pormenorizadamente la analogía entre las etiquetas de los archivos y los documentos que ponemos en ellos, con los nombres propios y las descripciones--definidas e indefinidas--, me gustaría analizar las siguientes dificultades. Tenemos un nombre propio y un conjunto de descripciones que asociamos con ese nombre. En algunas ocasiones, gracias a ese nombre y al conjunto asociado de descripciones, llegamos a creer que somos capaces de seleccionar un único objeto, el objeto que se ajusta precisamente a esas descripciones. Con el uso del nombre nos referimos a él. También creemos que el objeto que se ajusta a esas descripciones y el objeto al que tenemos la intención de referirnos cuando usamos esas descripciones y el nombre

⁴ El uso de esta analogía puede verse en Grice (1969) y Lockwood (1971).

propio que asociamos con ellas, es el mismo objeto único. Estas son las suposiciones básicas que sostenemos cada vez que usamos un nombre propio en un "discurso de actualidad". Sin embargo, como he sugerido anteriormente, puede suceder que (1) el objeto que, de hecho, se ajusta a esas descripciones no sea el mismo objeto al que intentamos referirnos cuando usamos aquellas descripciones. Puede suceder también que (2) dos o más objetos diferentes se ajusten de hecho a aquellas descripciones pero que nosotros sólo tengamos la intención de referirnos a uno de ellos porque, por ejemplo, sólo creamos que existe uno de tales objetos. Puede darse también el caso de que (3) el conjunto de descripciones que utilizamos no seleccione ningún objeto en absoluto, es decir, que no exista ningún objeto que satisfaga de hecho esas descripciones. De acuerdo con esta idea, y como muy bien ha señalado Donnellan al exponer el principio de las descripciones identificadoras (1972, 356), la función de estas expresiones es doble. Por una parte, ellas sirven como criterios para identificar el referente de los nombres y, por consiguiente, para determinar qué objeto o qué persona es el referente del nombre; por otra, sirven también para decidir que no hay tal referente (véase también Salmon (1981)).

En unión con esta triple distinción, merece la pena señalar que una de las ideas principales de Donnellan y Kripke en relación a la teoría de la referencia de los nombres propios--la que ha llegado a llamarse genéricamente "teoría de la referencia directa"--, es que las descripciones definidas que asociamos con los usos de un nombre pueden tener, al menos, dos orígenes bien definidos. Tenemos al menos dos fuentes diferentes para generar descripciones definidas e indefinidas dentro de nuestro sistema cognitivo, y, por consiguiente, para introducir nombres propios en nuestro vocabulario particular. Intuitivamente, las descripciones definidas pueden ser o informes verbales de experiencias visuales y de otro

tipo, o descripciones que recibimos a partir de los informes verbales de otros hablantes o de teorías más o menos bien trabadas. Llamaré al origen de la primera clase de descripciones "origen privilegiado" de tales descripciones, y al origen de la segunda clase de descripciones "origen parásito" tal y como Donnellan sugiere (1972, 373). Respecto de las descripciones de origen parásito dice este autor:

En general, nuestro uso de los nombres propios de personajes históricos (y también de los nombres de aquellas personas con las que no estamos familiarizados personalmente), es parásito de los usos de los nombres por otras personas--en conversación, en documentos escritos, etc. En tanto que poseamos un conjunto de descripciones definidas en estos casos, esas descripciones provienen de las cosas que se dijeron del presunto referente por otras personas. (1972, 373).

Por consiguiente, según la terminología previa, una descripción que sólo tiene un origen parásito es siempre una representación simbólica de su objeto. Una descripción que tiene un origen privilegiado en la percepción de su objeto es una representación propia de ese objeto. Esta idea, sin embargo, no debería llevarnos a pensar que los nombres propios se introducen necesariamente en nuestro vocabulario personal en razón exclusivamente de un conjunto de descripciones definidas que tienen su origen en lo que hemos oído hablar a otras personas acerca del objeto. Un nombre se puede introducir "a través" de la percepción del objeto mismo y *a fortiori* por ostensión, como Kripke señala. En estos dos casos, el usuario de un nombre propio tendrá acceso privilegiado al origen o fuente de un nombre propio, y tendrá éxito incondicionado a la hora de referirse al objeto al que aplica el nombre propio. Por consiguiente, la distinción primordial entre el origen parásito y el origen privilegiado de una descripción también puede expresarse con estas palabras: una descripción tiene un origen privilegiado para un hablante cuando éste está en

condiciones, o estuvo en condiciones en el momento de "generar" la descripción, de determinar y seleccionar el referente de una forma *independiente* de la que pueda proporcionar el contenido semántico de la descripción. Por ejemplo, la percepción de un objeto sería una de estas maneras independientes y privilegiadas de seleccionar la referencia.

& 16. Origen, cualidad y verdad de las descripciones.

Después de esta breve consideración acerca de los posibles orígenes de las descripciones que asociamos con nuestro uso de los nombres propios, es importante saber a qué clase de cosas nos referimos normalmente cuando usamos descripciones definidas. Básicamente y según la mayoría de los ejemplos que aparecen en la literatura sobre este tema, nos referimos a objetos con propiedades monádicas o diádicas. Estas propiedades, a su vez, pueden ser exclusivas de un objeto, pertenecer a diferentes objetos, o no corresponder a ningún objeto en absoluto. "Un orador romano" es, muy probablemente, una descripción a la que se ajustaba más de un ciudadano de aquel famoso imperio, incluido Cicerón. "El padre del autor de *Waverley*" se ajusta, por el contrario, a un único objeto, caso de que exista el autor de *Waverley*. "El caballo con setecientas veintiuna cabezas" es, probablemente, un ejemplo del tercer tipo—en nuestro mundo al menos, como suele decirse.

Además, toda descripción puede ser, básicamente, verdadera o falsa del objeto al que tenemos la intención de referirnos cuando usamos aquellas descripciones y el nombre propio asociado con ellas. Por ejemplo, supongamos que el archivo que etiqueto en mi sistema cognitivo como "Aristóteles" contiene las descripciones "el maestro de Alejandro Magno", "el alumno más famoso de Platón" y "el fundador de la escuela peripatética" entre los documentos allí almacenados. Pudiera suceder, sin

embargo, que la persona a la que intento referirme cuando uso el nombre propio "Aristóteles" y sus descripciones asociadas no se ajuste, de hecho, a todas estas descripciones. Digamos que tal persona no se ajusta a la descripción "el maestro de Alejandro Magno". Como Donnellan señala con su ejemplo de Tales y aparece también en el ejemplo similar de Gödel-Schmidt que ofrece Kripke (1980, 83-84), podemos descubrir, siquiera sea ficticiamente aunque no contradictoriamente, que la persona de la que quiero hablar cuando uso el nombre "Aristóteles" y sus descripciones asociadas no fuera de hecho el maestro de Alejandro Magno. En este caso, "el maestro de Alejandro Magno" sería una descripción falsa del objeto al que intento referirme cuando uso el nombre "Aristóteles". Si fuera verdad que llegara a saber que "el maestro de Alejandro Magno" es una descripción falsa de Aristóteles, eliminaría sin contemplaciones el documento "el maestro de Alejandro Magno" de mi archivo "Aristóteles". Sin embargo, es posible que "el maestro de Alejandro Magno" fuera una descripción verdadera de algún *otro* objeto: el objeto que fue de hecho el maestro de Alejandro Magno. Imagínese, por lo tanto, que llegáramos a saber que Aristóteles no fue el maestro de Alejandro Magno. Un experto en Historia Griega ha descubierto que un seguidor entusiasta de la filosofía de Aristóteles, del que hasta ahora no se tenía noticia especial alguna, llamado "Midias", fue de hecho el maestro de Alejandro Magno. Aristóteles estaba muy ocupado y pagó a Midias para que hiciera humildemente el trabajo. En un caso semejante, este experto en Historia Griega, aparte de su archivo "Aristóteles" con las descripciones "el alumno más famoso de Platón" y "el fundador de la escuela peripatética" entre sus documentos, tendría también el archivo "Midias" con su correspondiente descripción asociada: "el maestro de Alejandro Magno". El resultado de esta situación es que uno y el mismo documento--"el maestro de Alejandro Magno"--puede estar almacenado en diferentes archivos en diferentes sujetos. Dependiendo

del archivo en el que esté almacenada la descripción, el uso de la descripción hará que ésta sea verdadera o falsa del objeto al que los usuarios de diferentes archivos-nombres tengan la intención de referirse. En el descubrimiento ficticio propuesto, "el maestro de Alejandro Magno" es una descripción falsa del objeto al que un usuario del archivo "Aristóteles" tenga la intención de referirse con el uso de ese nombre, dado que es una descripción verdadera del objeto al que el experto de historia griega, apoyado en sus investigaciones, está intentando referirse cuando usa el archivo que ha bautizado recientemente con el nombre de "Midias"⁵.

Aquellos que, como Donnellan y Kripke, han defendido una teoría de la referencia directa han considerado que lo que tiene ésta de novedoso respecto de la teoría ortodoxa de los nombres propios y las descripciones definidas, es precisamente que el referente del nombre propio usado por una persona no tiene por qué coincidir necesariamente con el objeto que satisface de hecho la descripción o descripciones, o un número suficiente de ellas, que el hablante estaría dispuesto a asociar con ese nombre y que utiliza como criterio identificador del referente. Y eso en el mejor de los casos. Nada impide que las descripciones que el hablante asocia con su uso de un nombre propio sean, de hecho, todas falsas. En el ejemplo ficticio propuesto, la descripción "el maestro de Alejandro Magno" que yo asocio con el nombre "Aristóteles" es, de hecho, falsa del individuo Aristóteles. Lo mismo podría pasarle al resto de descripciones que estoy dispuesto a asociar con mi uso del nombre propio "Aristóteles". A pesar de todo, cuando uso el nombre "Aristóteles" estoy refiriéndome a un individuo, Aristóteles. Por consiguiente, poco importa que un hablante describa

⁵ De los argumentos basados en esta clase de ejemplos, de los argumentos semánticos, ha dicho Salmon que son los "más fuertes y persuasivos" de cuantos pueden ofrecerse para defender una teoría de la referencia directa de los términos singulares (1981, 29). Salmon distingue en este sentido entre argumentos modales, argumentos epistemológicos y argumentos semánticos (1981, 23-31).

erróneamente el objeto al que quiere referirse cuando usa un nombre para que, de hecho, se refiera a él. Lo único que parece imprescindible para que se dé de hecho la referencia, aparte de la intención del hablante, es algo que Kripke ha puesto muy claramente de manifiesto:

Un "bautismo" inicial tiene lugar. Aquí el objeto puede ser nombrado por ostensión, o el referente del nombre puede ser fijado por descripción. Cuando el nombre es "pasado de eslabón a eslabón" quien recibe el nombre debe intentar, pienso, cuando aprende el nombre, usarlo con la misma referencia que el hombre de quien lo ha oído. (1980, 96).

Por consiguiente y según la teoría de la referencia directa, lo verdaderamente importante para que el nombre usado por un hablante se refiera a su objeto—para que "N" se refiera, pues, a *N*—, es que ese nombre tenga su origen en la ceremonia iniciática de darle nombre cuando el objeto mismo estaba "ahí delante". Después el nombre va pasando de un hablante a otro y todos ellos quedan vinculados referencialmente al objeto en virtud de esta cadena causal continua de transmisión. Poco parece importar que a lo largo de la cadena causal cada hablante vaya quitando o poniendo descripciones al uso del nombre. Poco importa que estas descripciones estén al final completamente desvirtuadas respecto de las propiedades reales del objeto al que se refiere el nombre propio así adquirido. Lo que fija la referencia es que cada usuario del nombre está unido causalmente a otro usuario del mismo nombre, hasta llegar al usuario que dio nombre al objeto de una manera privilegiada y, en cierto sentido, hasta llegar también al objeto mismo. Cuando el uso de un término singular queda así encadenado a su referente se dice de él que es un *designador rígido* de ese objeto.

De acuerdo con la imagen de los archivos, es posible explicar sencillamente tres ejemplos diferentes. El primero de ellos es el caso en el

que existe fracaso en la denotación. Hay fracaso en la denotación cuando deberíamos haber eliminado todos y cada uno de los documentos que hemos puesto en un archivo, dado que es verdad que, por no existir semejante objeto, ninguno de estos documentos se ajusta al único objeto al que tratamos de referirnos cuando usamos la etiqueta de ese archivo y sus descripciones asociadas, aun cuando tales descripciones podrían ajustarse a otros objetos, bien en conjunto, bien por separado. Ese archivo y sus documentos, cuando menos, no pueden utilizarse al unísono y conjuntamente para mencionar nada. Existe la intención, el querer referirse a un objeto, pero no existe la denotación o referencia.

Desde luego, si algo pone de manifiesto el fracaso en la denotación es lo fluida y sutil que puede llegar a ser nuestra vida cognitiva. Como sugiere Fodor, nuestra vida cognitiva podría parecerse muchísimo, siquiera en cierto sentido, a una red asociativa sumamente compleja en la que los elementos asociados son, como mínimo, un número inmenso de enunciados (mentales) agrupados y diferenciados de determinadas maneras (Fodor 1987, 123).

A través de esta caracterización a vuela pluma de nuestra vida mental, se puede entender de una manera precisa que no toda frase denotante está por un objeto, como dice el Meinong de Russell. El archivo "Centauro" no está por ningún objeto (intencional), sino por una agrupación determinada de expresiones y oraciones que archivamos en él. No darse cuenta de este hecho ha originado que numerosos filósofos se hayan preocupado excesiva y, desde mi punto de vista, innecesariamente, por problemas que surgen exclusivamente cuando se capitula ante el espejismo de creer que se está mencionando un objeto cuando se usa el nombre "Centauro", y que esa mención "alcanza" un objeto (intencional), cuando en realidad se está simplemente ante un mero nodo de la malla asociativa que constituye nuestra vida mental, ante un nodo con la etiqueta

"Centauro" en el que hay guardados ciertos documentos. El referente de "Centauro", por decirlo en muy pocas palabras, no es un objeto (intencional). Es, más bien, un nodo determinado, y solidificado así por muy diversos motivos, de esa red asociativa, un nodo en que se quedan incrustados diversos enunciados mentales que, como en los ejemplos previamente trabajados, bien pueden ser ciertas descripciones definidas.

El segundo caso que se puede explicar con la metáfora de los archivos es el siguiente. Dos hablantes pueden tener la intención de referirse a uno y el mismo objeto, pero el archivo del primer hablante contiene más archivos de los que están contenidos en el archivo del segundo hablante. Sin embargo, no es necesariamente verdadero que el hablante que usa el archivo que contiene más documentos pueda utilizarlos para seleccionar con más certeza el único objeto al que él tiene la intención de referirse. En esta caso, sin duda alguna, ambos hablantes han de tener diferentes criterios para identificar la denotación. Por ejemplo, pensemos que el primer hablante tiene estas tres descripciones en su archivo "Aristóteles": "el alumno de Platón", "el maestro de Alejandro Magno" y "el fundador de la escuela peripatética". El segundo hablante, tiene las descripciones: "el maestro de Alejandro Magno", "el autor de la *República*", "el alumno de Protágoras" y "el hombre que tenía tres gatos" en el mismo archivo "Aristóteles". Ahora, supóngase de nuevo que el experto en Historia Griega descubre que Aristóteles no fue el maestro de Alejandro Magno. En este caso, a pesar de todo, el primer hablante quizás estuviera todavía en condiciones de identificar a Aristóteles, dado que Aristóteles se ajusta a las otras dos descripciones que él tiene en su archivo. Por el contrario, el segundo hablante tendrá bastantes apuros a la hora de identificar a Aristóteles, porque las restantes descripciones de su archivo no son verdaderas de Aristóteles, ni seleccionan un único objeto. Como

argumentaré más pormenorizadamente en otro capítulo, para hacer semántica es necesario suponer que alguien está anclado en la verdad.

Donnellan, por su parte, señala el tercer caso que quiero examinar. Si dos hablantes--un estudiante y un profesor--tienen diferentes conjuntos de descripciones identificadoras para seleccionar un objeto, digamos Aristóteles, y "cada uno afirma la existencia de Aristóteles, hay al menos una posibilidad teórica de que uno <el profesor> esté en lo cierto, y el otro equivocado" (Donnellan 1972, 362). En este caso, el profesor y el alumno han de tener, de nuevo, diferentes criterios para identificar la denotación y para establecer su existencia. El profesor sabe más que su alumno acerca de Aristóteles, de tal forma que puede identificar a Aristóteles y determinar su existencia con más elementos de juicio y con más certeza que su alumno. Esto, por su parte, sólo puede suceder bajo una de estas dos condiciones:

(i) El conjunto de descripciones que el profesor etiqueta como "Aristóteles" no tiene como subconjunto las mismas descripciones que el alumno etiqueta también como "Aristóteles".

(ii) Si el conjunto de descripciones del profesor contiene las descripciones del estudiante como uno de sus subconjuntos, entonces entre las descripciones que el profesor tiene, pero que no tiene el alumno, ha de haber algunas que sean verdaderas de Aristóteles y sean suficientes para seleccionar este objeto.

Corolario: puede suceder que tengamos la intención de referirnos a un único objeto con un nombre propio y con sus descripciones asociadas; que sólo parte de estas descripciones sean verdaderas de este objeto y que, sin embargo, esta porción de descripciones que son verdaderas no sea suficiente para seleccionar un único objeto. Puedo archivar en "Aristóteles" las descripciones "el hombre con una cabeza", "el hombre con dos pies", "el asesino de Alejandro Magno" y no seleccionar ningún objeto si es que,

por otros caminos, no tengo ningún acceso privilegiado a este objeto, o alguna clase de contacto epistémico o causal con él, cosa ya bastante difícil de tener.

De acuerdo con este modelo de los archivos que he presentado, estoy de acuerdo con la primera condición del principio de las descripciones definidas tal y como lo expresa Donnellan:

El usuario o usuarios de un nombre propio debe ser capaz de suministrar un conjunto de, como las llamaré, descripciones "que no cometan una petición de principio" en contestación a la pregunta "¿a quién (o a qué) se refiere el nombre?" (1972, 360).

Esta condición sólo se puede sostener si entre el número "suficiente" de descripciones definidas hay al menos, o bien una descripción verdadera que expresa una propiedad que semejante objeto posee exclusivamente, o estas descripciones valen todas ellas juntas como una propiedad que semejante objeto posee exclusivamente. Estoy de acuerdo con la idea de que los documentos que se almacenan en un archivo puede que no sean suficientes--desde el punto de vista de un observador externo--para seleccionar el objeto al que el hablante está tratando de referirse. Por ejemplo, las descripciones del hablante pueden seleccionar dos o más objetos.

& 17. Nuevos ejemplos.

El ejemplo de Donnellan del niño que informa a sus padres de que "Tom es un hombre estupendo" a la mañana siguiente de una fiesta, tiene un interés especial cuando se tiene el propósito de explicar el rompecabezas que propone Russell. El archivo "Tom" que posee el niño en su cabeza tiene estos documentos: "el hombre estupendo que estuvo en mi habitación

la última noche", "el hombre que estuvo en la fiesta la última noche" y, quizás, "el hombre que me contó un cuento". Con estas descripciones a la mano el niño es incapaz, sin embargo, de reconocer otra vez al hombre que estuvo en su habitación la noche de la última fiesta. Imagínese que se olvidó de su aspecto porque la iluminación era muy débil y estaba medio dormido cuando Tom entró en su habitación y le contó un cuento. Sus padres, además, dan numerosas fiestas a las que acuden normalmente varios hombres que llevan ese mismo nombre como para poder decidir, sobre la base exclusiva de las descripciones que da su hijo, a cuál de ellos se está refiriendo el niño. Lo que es importante destacar en este caso es que, desde el punto de vista de sus padres, ellos también son incapaces de seleccionar un único objeto que se ajuste al conjunto de descripciones que da el niño si es que no tienen otras fuentes de información a las que puedan acudir. Por supuesto, la imposibilidad de seleccionar de nuevo al hombre con el que estuvo hablando el niño la última noche no significa que no haya una persona a la que el niño estaba tratando de referirse de una forma única cuando usaba el nombre "Tom" (Donnellan 1972, 364). De hecho, los padres del niño creen firmemente que él está intentando referirse a alguien. ¿A qué alguien? Como muy bien dice Brian Loar en su comentario a este ejemplo de Donnellan, los padres piensan que el niño "se está refiriendo al Tom que fue la causa de su uso de "Tom" " (1976, 367). La razón de esta afirmación es que el origen de las descripciones definidas del niño se supone que son diferentes percepciones de un hombre que estuvo en su habitación la última noche y que le contó un cuento. El niño usa sus descripciones definidas para informar acerca de lo que él cree que vio la última noche. En este caso, aunque el conjunto de descripciones que el niño da no sea suficiente para seleccionar un único objeto, y eso tanto desde el punto de vista de la primera persona-del niño--, como del de la tercera-de los padres--, el niño, de todas formas,

estaba intentando referirse a un único objeto, porque el origen de sus descripciones y del uso del nombre en que las almacenó era un supuesta percepción de un cierto hombre, de *ese* Tom. La descripción "el hombre agradable que estuvo en mi habitación la última noche" tiene, pues, un origen privilegiado para el niño. Los padres simplemente le creen.

Si este es un ejemplo razonable, parece que la cuestión de a qué se refiere un hablante cuando usa un nombre no se excluye por su incapacidad para describir o incluso reconocer o señalar al referente. (Donnellan 1972, 364).

Esta tesis es muy plausible y, una vez más, yo estoy de acuerdo con ella. Piénsese en lo que sucede cuando la policía comienza la investigación oficial de un asesinato sobre el dato de que hay una persona muerta en la calle y de que un testigo informa, mediante algunas pocas descripciones, acerca del asesino que apenas pudo ver. Aunque las descripciones del testigo fundadas sobre una percepción no seleccionen de por sí un objeto único, finalmente se utilizan como el punto de partida de la investigación policial que, si fructífera, dará finalmente con el asesino.

El análisis de otro ejemplo de Donnellan, el de los dos cuadrados dibujados en una pantalla, proporciona nuevos motivos de reflexión acerca de qué es aquello que está verdaderamente implicado en la referencia de los términos singulares. Un sujeto sometido a un experimento psicológico ve dos cuadrados iguales frente a sí, y se le propone la tarea de nombrarlos y de explicar qué criterio utiliza para nombrarlos de esa manera. El cuadrado que este sujeto ve en la parte de arriba de la pantalla es llamado por él "alpha", y el cuadrado que ve en la parte de abajo es llamado por él "beta". Según el modelo que aquí se está proponiendo, el hablante ha abierto ante esta nueva situación dos nuevos archivos en su sistema cognitivo que etiqueta, respectivamente, como "alpha" y "beta". En "alpha"

ha puesto la descripción "el cuadrado de arriba", y en "beta" la descripción "el cuadrado de abajo". Las descripciones "el cuadrado de arriba" y "el cuadrado de abajo" son las únicas descripciones identificadoras que él puede dar de los objetos a los que se refieren respectivamente los nombres "alpha" y "beta". Por supuesto, el hablante piensa que el objeto que se ajusta *verdaderamente* a aquellas descripciones es el mismo objeto al que él tiene la intención de referirse mediante el uso de los nombres "alpha" y "beta". O, de otra manera, que el hablante cree que algo que en verdad está arriba ha sido la causa de las propiedades de su experiencia visual y que él describe brevemente como "alpha" y "beta". Desde el punto de vista del hablante, el origen de sus descripciones es una percepción verdadera de dos cuadrados y de su disposición espacial relativa. Es decir, sus descripciones tienen para él un origen privilegiado. Pero ahora aparece una dificultad inesperada. El hablante lleva puestas unas gafas especiales cuyo curioso efecto es que ve en la parte superior de su campo visual lo que verdaderamente está en la parte inferior, y viceversa. El experimentador, sin embargo, *sabe* el efecto que obran las gafas en las experiencias visuales del hablante y, como consecuencia, el efecto de las gafas en los informes verbales de estas experiencias. El experimentador sabe cuál es la situación real. El experimentador sabe que lo que el hablante ve en la parte de arriba está realmente en la parte de abajo, y viceversa. Por consiguiente, el experimentador tiene dos archivos, que él también llama "alpha" y "beta", cuyas descripciones tienen origen en una percepción verdadera de la situación espacial relativa de los cuadrados. El archivo "alpha" del experimentador contendría, según la situación descrita, el documento "el cuadrado que el hablante ve en la parte superior de la pantalla y que él llama "alpha" está realmente en la parte inferior de la pantalla". Correlativamente, su archivo "beta" contiene el documento "el cuadrado que el hablante ve en la parte inferior de la pantalla y que llama "beta" está

realmente en la parte superior de la pantalla". Así, pues, "alpha" para el experimentador es el cuadrado de abajo y "beta" el cuadrado de arriba, justo al contrario que para el sujeto de experimentación.

Ahora cambiamos el color del cuadrado que está realmente abajo y preguntamos al sujeto qué cuadrado ha cambiado de color. El sujeto, como puede fácilmente imaginarse, contestará que "alpha", esto es, "el cuadrado de arriba es el que ha cambiado de color". La pregunta es: ¿a qué cuadrado está refiriéndose ahora el sujeto cuando dice que ha sido "alpha" el cuadrado que ha cambiado de color? En relación con este ejemplo, Donnellan comenta:

... pero dado *nuestro conocimiento* de la presencia y del efecto de las gafas invertidoras y de la *ignorancia del sujeto* en relación a esto, parece claro que deberíamos tomarle como refiriéndose, no al cuadrado de arriba, sino al que le parece a él *erróneamente* que está arriba-el que está abajo... Creo que sería del todo correcto decir que aunque *él no lo sepa*, está hablando acerca del cuadrado que está abajo, aunque él lo *describiera* como "el cuadrado de arriba" (1972, 368-369; los subrayados son míos).

El problema de este texto es saber qué entiende Donnellan cuando dice que "el <hablante> está hablando acerca del cuadrado de abajo" cuando, como contestación a la pregunta "¿qué cuadrado ha cambiado de color?", dice que alpha es el cuadrado que ha cambiado de color.

Estoy de acuerdo con esta afirmación si Donnellan quiere insinuar con ella que el hablante pondría el nombre de "alpha" al cuadrado de la parte inferior, como hace el experimentador, si tuviera un verdadero conocimiento de la situación real, y que sería de este cuadrado del que diría que ha cambiado de color. Desde el punto de vista del experimentador, cuando el hablante dice "el cuadrado de arriba", el experimentador "sabe" que si el hablante supiera cuál es la situación real, el hablante informaría de su experiencia visual diciendo que el cuadrado de

abajo es el que ha cambiado de color. En otras palabras, si el hablante conociera la situación de hecho, intercambiaría las etiquetas "alpha" y "beta" entre sus archivos.

Pero no estoy de acuerdo con Donnellan si es que quiere dar a entender que cuando el hablante dice "el cuadrado de arriba", tiene la intención de referirse al cuadrado que realmente está en la parte inferior. Desde el punto de vista del hablante, él está intentando referirse al cuadrado que *él cree* que está arriba. Desde el punto de vista del experimentador, por ser un conocedor ideal, el hablante está de hecho refiriéndose a la disposición espacial con que se le ofrecen ciertos individuos dentro de su campo visual. Para el hablante, su informe verbal lo sería, así lo cree él, acerca de un objeto real cuadrado que está en la parte superior de la pantalla y que ha cambiado de color. En virtud de la estructura de flecha de sus datos de los sentidos junto con sus aprehensiones respectivas, él está refiriéndose (simbólicamente) al cuadrado que él cree que está en la parte de arriba. Para el experimentador, el informe del hablante habla acerca de la posición espacial relativa de algo acerca de lo cual el experimentador "que sabe" caracterizaría como los datos de los sentidos del hablante. Simplemente, el cuadrado que el hablante dice que está arriba, está de hecho abajo, y viceversa.

Por supuesto, no quiero dar a entender que sea falso decir que el hablante podría estar refiriéndose a algo distinto de lo que él cree que está refiriéndose. De hecho, esto es lo que sucede realmente. El hablante probablemente está refiriéndose al cuadrado que está en la parte de abajo cuando él dice la oración: "*alpha* es el cuadrado que ha cambiado de color". Sin embargo, esta afirmación sólo la puede sostener un observador externo que está en una relación cognitiva perceptual privilegiada con ambos cuadrados. Adviértase que no definiendo que alguien esté refiriéndose a algo distinto de aquello a lo que él cree que está refiriéndose cuando y

solamente cuando un observador externo sea capaz de informar de la situación real. Lo que estoy sugiriendo es que los seres humanos sólo pueden hacer justificadamente esta afirmación cuando tienen razones teóricas o empíricas fundadas para hacerlo así.

Podemos obtener nuevas intuiciones acerca de la relación entre los nombres propios, las descripciones definidas y el conocimiento analizando el ejemplo del filósofo ficticio Aston-Martin. Un entusiasta de la filosofía ha oído hablar de Aston-Martin como del ilustre filósofo autor de "*Otros cuerpos*" y máximo defensor de la teoría del egocentrismo pluralista. Ha leído incluso sus obras. Desgraciadamente, el hablante no tiene ninguna familiaridad directa con el objeto al que él está tratando de referirse cuando usa el nombre "Aston-Martin" y sus descripciones asociadas. Sólo sabe de él de oídas, parásitamente. Todo lo que él tiene es un archivo con la etiqueta "Aston-Martin" y un conjunto de descripciones asociadas, cuyo origen es enteramente parásito, y que él cree seleccionan un único objeto: Aston-Martin mismo. Dado que entiende esas descripciones, él tiene la intención de referirse a ese objeto cuando usa el nombre "Aston-Martin". Así, pues, para el hablante, el archivo "Aston-Martin" incluye estas descripciones: (1) "el autor de *Otros cuerpos*" y (2) "el expositor principal de la teoría del egocentrismo pluralista".

Con estas herramientas para seleccionar objetos, el hablante va muy ufano un día a una fiesta en la que tiene la esperanza de encontrarse y hablar con tan famoso filósofo. De hecho, el hablante se encuentra ya en la fiesta y conoce a un hombre que él cree que es el mismo Aston-Martin al que trataba de referirse antes cuando usaba el archivo con la etiqueta "Aston-Martin".

El siguiente día, el hablante cuenta a sus amigos, todo contento, que la última noche estuvo hablando con Aston-Martin, el famoso filósofo que escribió "*Otros cuerpos*" y que es el fundador de la teoría del egocentrismo

pluralista. En el tiempo en que está hablando con sus amigos, probablemente, ya ha añadido nuevos documentos en su archivo "Aston-Martin" en relación a los que tenía previamente. Todo lo que sucedió durante la fiesta es seguramente motivo más que suficiente para incorporar nuevas descripciones en su archivo "Aston-Martin". Dado que durante la fiesta tuvo ocasión de hablar con el hombre que él creía que era Aston-Martin y de jugar con él a un juego de geografía que versaba sobre ciudades ahora tiene, además de las descripciones (1) y (2), las descripciones (3): "el hombre con el que estuve hablando ayer en la fiesta" y (4): "el hombre con quien jugué un juego de geografía que versaba sobre ciudades".

El único problema en este punto es que el hablante estuvo hablando con un hombre que no se ajustaba de hecho a las descripciones (1) y (2). Es decir, que no estuvo hablando con Aston-Martin el filósofo. Sin embargo, es porque el hablante creía que sí que estuvo hablando con un hombre que se ajustaba a las descripciones (1) y (2) por lo que fue capaz de añadir después las descripciones (3) y (4) a su archivo "Aston-Martin". De hecho hay un amigo del hablante que sabe que su amigo no conoció en realidad a Aston-Martin, el filósofo, sino a otro hombre con el mismo nombre a quien no le gusta defraudar las expectativas de los demás. Su amigo, que está enterado de lo que pasa, tiene dos archivos con la misma etiqueta "Aston-Martin". Probablemente él "marque" estas etiquetas de tal modo que pueda distinguir, por ejemplo, entre "Aston-Martin=" y "Aston-Martin*". En "Aston-Martin=" almacena los documentos (1) y (2). En "Aston-Martin*" almacena los documentos (3) y (4).

El amigo que está enterado dice a su amigo y a la audiencia que "el hombre con quien el hablante estuvo hablando la última noche no era Aston-Martin= con las propiedades que se expresan en (1) y en (2), sino Aston-Martin* con las propiedades expresadas en (3) y (4), y esto es verdad". Automáticamente, el hablante y la audiencia crearán dos nuevos

archivos. "Aston-Martin=" con los documentos (1) y (2) y "Aston-Martin*" con los documentos (3) y (4), si es que le creen. En este caso, todo el mundo está enterado y afirmará que cuando el hablante dijo hace un momento "yo estuve hablando con Aston-Martin y estuvimos jugando a un juego de geografía", él tenía la intención de referirse a Aston-Martin el filósofo pero que *fracasó* al hacerlo. Contrariamente a lo que Donnellan parece sugerir, no es del todo exacto decir que él tenía la intención de referirse a "Aston-Martin"—en el sentido de "Aston-Martin=" y de sus documentos asociados (1) y (2)—, sino que se refirió a "el hombre con el que estuvo hablando la última noche y con el que estuvo jugando a un juego de geografía sobre ciudades"—en el sentido de "Aston-Martin*" y de sus documentos (3) y (4) (Donnellan 1972, 369).

Desde el punto de vista del hablante, cuando él no sabía cuál era la situación real, esto es, antes de saber que debería haber tenido dos archivos en donde solamente tenía uno, el referente de "el autor de *Otros cuerpos*" y el referente de "el hombre con el que estuve hablando ayer", es el mismo.

En resumen, en el ejemplo de los cuadrados y en el ejemplo de Aston-Martin, hay siempre una situación "antes del conocimiento" y una situación "después del conocimiento". En la situación "antes del conocimiento" puede suceder que el objeto que el hablante cree que selecciona la descripción de una manera única no sea el objeto que hubiera seleccionado en la situación "después del conocimiento".

A la luz de todos los ejemplos analizados no creo que la teoría de Donnellan, a la que él llama teoría histórica o teoría de los bloques, y la teoría de la cadena causal de transmisión de Kripke, sean alternativas a la teoría de las descripciones y de la creación de archivos⁶. Son sólo

⁶ Por este motivo no puedo estar de acuerdo con Searle cuando afirma que la teoría causal de los nombres propios es, en el fondo, equivalente a la de las descripciones definidas (1982).

teorías que ponen un énfasis especial en el origen de nuestros documentos en los archivos y, como consecuencia, en la importancia que tiene poseer un acceso privilegiado a la fuente de la descripción para estar seguros de que el referente de nuestros nombres y de nuestras descripciones definidas es, de hecho, el objeto que se ajusta a esas descripciones y al que se refiere el nombre. Incluso el propio Kripke da pie a pensar de esta manera. Así, por ejemplo, en una nota a pie de página muy mencionada dice lo siguiente:

Dos cosas se deberían subrayar concernientes al caso de la introducción de un nombre por medio de una descripción en un bautismo inicial. Primera, que la descripción que se usa no es sinónima con el nombre que ella introduce, sino que más bien fija su referencia. Aquí diferimos de los teóricos normales de las descripciones. Segunda, que muchos casos de bautismo inicial están muy lejos de aquellos que inicialmente inspiraron la teoría de las descripciones. *Usualmente el que bautiza está familiarizado en algún sentido con el objeto que nombra y es capaz de nombrarlo ostensivamente.* (1980, 96. El subrayado es mío).

& 18. La teoría de la referencia y sus dificultades.

La teoría de la referencia de los nombres propios que estoy proponiendo se puede condensar en esta formulación: cuando una persona usa un nombre "N" *no* se refiere necesariamente al objeto único que se ajusta a las descripciones que dicha persona asocia con "N", sino más bien al objeto único que es el *origen* de esas descripciones para esa persona. Si las descripciones que asocia con "N" tienen diferentes orígenes—como en el ejemplo de Aston-Martin—, entonces lo que importa son las descripciones que tienen un origen privilegiado. La referencia del nombre será el objeto único que es el origen de las descripciones *privilegiadas*.

En un principio la idea es ésta. Sin embargo, no conozco ninguna teoría de la referencia de los nombres propios que no tenga problemas. (¡Una cosa tan tonta!) La mía no iba a ser una excepción⁷. El principal de todos se puede condensar en la siguiente pregunta: ¿y qué pasa si alguien pone en el mismo archivo-nombre descripciones que tienen orígenes privilegiados diferentes? Supóngase que hay dos gemelas, Patty y Vicky. Un día me encuentro con Patty y al día siguiente me encuentro con Vicky a quien confundo con Patty (no me doy cuenta de que hay dos personas). En esas condiciones pongo en el mismo archivo-nombre "Patty" las descripciones que he adquirido después de ambos encuentros. Y bien, ¿a cuál de las dos me refiero ahora cuando uso el nombre "Patty"? Todas las descripciones tienen un origen privilegiado y, por lo tanto, no puedo usar el criterio de las descripciones para elegir entre las dos hermanas y fijar así la referencia.

Creo que mi teoría podría salir momentáneamente del paso de la siguiente manera. Desde tu punto de vista, dado que tú sólo tienes un archivo, te refieres sólo a una persona. Pero desde el punto de vista de las personas que *saben* cuál es la situación real—aquellas que mantienen archivos diferentes para Patty y para Vicky—tú te estás refiriendo a diferentes personas en diferentes ocasiones, dependiendo de qué descripciones sean relevantes en cada ocasión.

Sin embargo, después de todo, esta salida no es del todo tan buena, aunque capte de una manera adecuada las diferencias entre las creencias de las personas y el modo en el que asociamos descripciones con nombres. Pero todavía deja una parte del rompecabezas sin tocar. *¿A quién me estoy refiriendo de hecho con mi uso del nombre "Patty" en una ocasión determinada?* Decir a quién me estoy refiriendo desde el punto de vista de

⁷ Debo a la profesora Leora Weitzman de la Universidad de Wisconsin-Madison el plantamiento de todas estas dificultades.

alguien más no es contestar la pregunta. No vale hacer semántica *ex machina*. Incluso así, sigue siendo un poco desorientador decir que desde el punto de vista de alguien *que sabe* me estoy refiriendo sólo a una persona en cada ocasión, aunque yo no lo sepa. ¿A quién? ¿A Patty o a Vicky? No hay una única persona Patty-Vicky a quien pudiera estar refiriéndome por mucho que crea que sólo hay una persona que se ajusta a mis descripciones. Lo que ha sucedido en realidad en este caso es que he cometido un error a la hora de archivar las descripciones y pudiera ser que el resultado de este error fuera que *no puedo estar refiriéndome a nadie* con mi uso del nombre "Patty" hasta que no consiga enderezar las cosas en mi cabeza.

Como se verá en el capítulo siguiente, ésta es una dificultad con la que también se enfrentan algunas teorías semánticas que tienen por objeto estudiar las condiciones de verdad de las adscripciones de actitudes proposicionales. Por mi parte, reconozco la dificultad y reconozco también que no sé cómo resolverla. De todas formas, sí que creo que el esbozo de teoría presentado hasta aquí es suficientemente adecuado como para resolver el rompecabezas de Russell.

§ 19. La solución del rompecabezas de Russell.

Con todas las herramientas conceptuales y terminológicas propuestas hasta este momento se puede descifrar el rompecabezas de Russell de esta manera. Jorge IV no sabía que Scott era el autor de *Waverley*, aunque Russell tuviera el conocimiento verdadero de que Scott fue el autor de *Waverley*. Russell era, por lo tanto, capaz de sustituir el nombre "Scott" por la descripción "el autor de *Waverley*" en todos los enunciados en los que "Scott" aparece y el nombre "Scott" se refiere al mismo objeto que denota "el autor de *Waverley*". Y, seguramente, Jorge IV no estaba

especialmente interesado en la ley de identidad. Más bien, estaba interesado en saber si la persona a la que tiene la intención de referirse cuando él dice "Scott" es la misma persona que se denota comúnmente con el significado de la frase "el autor de *Waverley*".

Imaginemos un caso que provea una razón por la cual Jorge IV tuviera ya el archivo "Scott" en su sistema representacional. El puede saber, por ejemplo, que el autor de *Ivanhoe* fue Walter Scott. Lo que él quiere saber cuando pregunta si Scott era el autor de *Waverley* es, en el fondo, lo mismo que saber si el autor de *Ivanhoe* era el autor de *Waverley*. Russell, por su parte, que sabía muy bien que Scott fue el autor de *Waverley*, podría estar interesado en la ley de identidad, pero Jorge IV, quien no sabía que "Scott es el autor de *Waverley*" es un enunciado de identidad verdadero, no estaba interesado en tan interesante ley. Jorge IV sólo estaba interesado en tener el mismo conocimiento que tenía Russell acerca de Scott y del autor de *Waverley*; Jorge IV estaba interesado en tener la misma clase de conocimiento que permitía a Russell sustituir el nombre "Scott" por la descripción "el autor de *Waverley*" en toda proposición en la que uno de estos términos singulares apareciera y en poderlas reconocer como verdaderas. La contribución que las expresiones "Scott" y "el autor de *Waverley*" hacen a la proposición en la cabeza de Jorge IV es esencialmente diferente.

A esta línea de argumentación la profesora Leora Weitzman me sugirió de nuevo la siguiente dificultad interesantísima. Siempre hay pegas. De acuerdo con la teoría de la referencia que aquí propongo, lo que hace que el uso por parte de Russell de la descripción "el autor de *Waverley*" se refiera a Scott, es que dicho uso tiene su *origen* en Scott. Muy probablemente, el uso por parte de Jorge IV de la misma descripción tiene también su origen en Scott, tanto si Jorge IV lo sabe como si no. Y esto, además, podría haber sucedido, aunque ni Russell ni Jorge IV hubieran

obtenido dicha descripción de una manera privilegiada, sino sólo de una forma parásita. Por consiguiente, el uso por parte de Jorge IV de "el autor de *Waverley*" se refiere a Scott. Pero el uso por parte de Jorge IV del nombre "Scott" probablemente también se refiera a Scott, dado que las descripciones que ya asocia con "Scott" se originaron en Scott. Al menos, su uso de "Scott" y su uso de "el autor de *Waverley*" se *podrían* haber originado en la misma persona sin él saberlo. Pero entonces, cuando Jorge IV pregunta "¿es Scott el autor de *Waverley*?", está refiriéndose a la misma persona dos veces, y *si el significado de su pregunta está dado por las referencias de las expresiones* "Scott" y "el autor de *Waverley*", entonces parece que él ha preguntado, después de todo, por la ley de identidad.

Contra esta objeción podría señalar de momento que el significado completo de su pregunta tiene que ver con algo más que con la referencia o referencias de las expresiones que aparecen en ella. Tiene también algo que ver con los documentos que Jorge IV almacenaría en sus archivos.

En cualquier caso, aparte de esta objeción, hay otro problema de índole más general que se puede alegar en contra del esquema de teoría que he intentado presentar. La analogía de los archivos ilustra, pienso, de una manera clara la forma en la que nosotros, como *sistemas cognitivos*, podríamos funcionar, y cómo es que podemos cometer cierta clase de errores, albergar cierta clase de incertidumbres y aprender cierta clase de cosas. Pero no se encara completamente con las cuestiones *semánticas* que tienen que ver con aquello a lo que se refieren nuestros usos de los nombres en ciertas circunstancias que constituyen un rompecabezas lógico. ¿Cuál es la teoría completa acerca del significado del uso de la descripción "el autor de *Waverley*" por parte de Jorge IV?

Si se usa la analogía entre los nombres propios y las etiquetas de los archivos, la situación se puede explicar de la siguiente manera. Cuando Jorge IV quiso saber si Scott era el autor de *Waverley*, estaba interesado

en saber si debería engrosar su archivo "Walter Scott" con el documento "el autor de *Waverley*". A él, naturalmente, le gustaría obtener esta información de un hablante cuyos documentos en el archivo "Walter Scott" tuvieran un origen privilegiado y no meramente parásito. Cuando Russell afirma que "de hecho, Scott fue el autor de *Waverley*", podemos suponer que él estaba enterado de que el documento "el autor de *Waverley*" debería estar almacenado en el archivo "Walter Scott" de Jorge IV, en donde, quizá, el documento "el autor de *Ivanhoe*" estaba ya archivado de antes. De acuerdo con esta sugerencia, el texto de Russell que cité en el párrafo 14 se podría parafrasear de esta otra manera: si *a* es idéntico con *b*, lo que sea verdad del uno, lo será del otro, y "a" puede ser sustituido por "b" en cualquier proposición sin alterar la verdad o falsedad de la proposición. Ahora bien, Jorge IV quería saber si Scott fue el autor de *Waverley*; y, de hecho, Scott—cierto Scott, *ese* Scott—, fue el autor de *Waverley*. Es por este motivo por el que Russell podía sustituir la etiqueta "Scott" por uno de los documentos que él había almacenado en ese archivo, a saber, por "el autor de *Waverley*", en toda proposición en la que apareciese el nombre "Scott", y, de aquí, investigar si Jorge IV deseaba saber si el documento "el autor de *Waverley*" debería estar almacenado en su archivo "Walter Scott". Si esto no es así, entonces deberíamos atribuir un interés especial en la ley de identidad al que era el primer gentilhomme de Europa. Pero semejante interés, como bien señala Russell, no se le atribuye a Jorge IV.

En otras palabras, la circunstancia de que dos términos singulares sean estrictamente coextensivos no entraña que cualquier sistema cognitivo tuviera que reconocerlos inmediatamente como necesariamente coextensivos. Como consecuencia, alguien podría fracasar a la hora de reconocer como idénticas las proposiciones en las que aparecen estos términos singulares como partes semánticamente relevantes, es decir, como proposiciones que

se refieren a, o nombran, el mismo estado de cosas. Las proposiciones (1) "el autor de *Waverley* tiene un perro" y (2) "Walter Scott tiene un perro" son dos proposiciones que se refieren al mismo estado de cosas. Si el autor de *Waverley* es idéntico con Walter Scott, ambos son la misma persona y es esa persona la que tiene un perro. El resultado de este hecho es que los valores de verdad de las proposiciones en las que los términos singulares que denotan o se refieren a ese escritor resultan ser partes semánticamente relevantes, son los mismos valores de verdad cuando sustituimos un término singular por el otro. Sin embargo, este hecho no implica que un creyente tuviera que creer en la verdad de la proposición (1) cuando cree en la verdad de la proposición (2). Parte de la razón de por qué esto es así es porque los nombres propios son signos completamente arbitrarios o etiquetas sin ningún significado. Usamos estas etiquetas, que forman parte de un sistema de convenciones lingüísticas, para referirnos a las cosas particulares, y, como es obvio, no podemos conocer todas estas convenciones lingüísticas dadas nuestras evidentes limitaciones cognitivas. (Las cosas, por supuesto, no tienen nombre. Se los ponemos.) Sin embargo, si fuéramos sistemas cognitivos perfectos, tenderíamos seguramente a usar la misma etiqueta para referirnos a la misma cosa con el fin de evitar así innecesarias y embarazosas proliferaciones de términos singulares. Pero esto es sólo un ideal.

CAPITULO 4.

LA TEORIA DE LAS ACTITUDES PROPOSICIONALES DE MARK RICHARD.

& 1. Hacia una teoría semántica general de las actitudes proposicionales.

Una recapitulación muy breve de los tres capítulos previos es ésta: la adscripción de actitudes proposicionales a los organismos--la instancia intencional--, es un paso esencial para explicar la etiología de su conducta inteligente y propositiva. Estas actitudes proposicionales aparecen investidas, desde la reflexión sobre la psicología popular, con propiedades semánticas y causales. La pregunta fundamental a que dieron lugar todas las argumentaciones del capítulo primero fue la de saber cómo es posible que haya algo que tenga simultáneamente estas dos clases de propiedades. Pero la reflexión inaugural sobre el fundamento inteligible de la ciencia psicológica debía dar paso necesariamente a la investigación de su fundamento ontológico. Para intentar explicar este extrañísimo entrelazamiento de propiedades semánticas y causales que se da de una manera única en las actitudes proposicionales, es preciso postular símbolos o representaciones mentales de naturaleza últimamente material, símbolos a los que se adscribe, sin embargo, contenido semántico. Esta es la estrategia que estaría operando implícitamente en el fondo de una buena parte de la

labor psicológica cognitiva actual. Sin excepción alguna, todas las páginas del capítulo segundo estuvieron dedicadas a la investigación del problema ontológico de las actitudes proposicionales, a la investigación de su naturaleza "anómala".

En el capítulo tercero comenzaron a cambiar las tornas. Con el ejemplo de don Quijote y Sancho se empezó a ver de una forma más concreta el modo en que realizamos adscripciones de actitudes proposicionales a los sujetos--en concreto atribuciones de estados perceptivos y alucinatorios--, así como a analizar los supuestos de los que dependen estas adscripciones. Dicha investigación condujo progresivamente al descubrimiento de que el problema de las actitudes proposicionales y el de su adscripción no podían aislarse enteramente de otros problemas filosóficos quizás más amplios y radicales. Hacer de un ser determinado el sujeto de ciertas predicaciones psicológicas esenciales--de estados perceptivos, por ejemplo--y hacer esto justificadamente, implica hablar de que sus estados psicológicos están relacionados con el mundo de los objetos físicos. El análisis de la percepción, en tanto que fenómeno psicológico básico gracias al cual conocemos el mundo, puso inmediatamente de manifiesto que el problema aparentemente particular de la adscripción justificada de actitudes proposicionales a un organismo abre expectativas y nuevos problemas directamente relevantes para la teoría del conocimiento y la metafísica en su sentido más "clásico" e interesante. El mundo del sujeto no sólo es mundo psicológico, sino mundo de conocimiento del mundo.

A partir de aquí la multiplicación de las direcciones en las que debe adentrarse el filósofo parece no conocer fin. Inmediatamente surgió la dicotomía entre realismo e idealismo a propósito de la tesis generalísima de la existencia del mundo. La percepción (verdadera) es siempre percepción de un trozo de mundo realmente existente y conocido. Esta pieza me parece fundamental y su resolución en uno u otro sentido completamente

definitiva dentro de la argumentación general. Solamente sobre la tesis del realismo metafísico es posible asentar fundadamente la noción de una ciencia psicológica natural para la cual el organismo, su objeto de estudio por excelencia, es parte natural de un mundo natural perfectamente real. Un ser que es sujeto de atribuciones psicológicas es un ser que ha de desenvolverse de una manera organizada e inteligente en el mundo a lo largo del tiempo.

Desde este momento y explotando problemas que están involucrados en el análisis de la percepción, fue surgiendo poco a poco el problema de la referencia. Con esta forma de proceder se consiguió asentar la teoría semántica y ontológica de las actitudes proposicionales en su contexto más amplio y radical: la teoría del conocimiento y la ontología general. En concreto, el deseo de aclarar algunas dificultades básicas en relación a los factores que intervienen en las percepciones y las alucinaciones, como dos fenómenos esencialísimos de los que se ha ocupado clásicamente la psicología de todos los tiempos, obligó a ocuparse de la metafísica. Con este punto de apoyo se empezó a delinear la teoría semántica de los nombres propios y de las descripciones definidas. Estas dos clases de expresiones son dos de las formas más elementales con las que contamos para referirnos a lo que hay, a los objetos del mundo, y a sus propiedades.

Este trabajo previo es realmente esencial, como se podrá comprobar constantemente a lo largo de este capítulo dedicado a estudiar las propiedades semánticas y estructurales de las representaciones mentales. Sin una teoría mínima de la referencia y, en especial, sin una teoría de la referencia de los nombres propios y las descripciones definidas, es imposible elaborar una teoría de las proposiciones, es decir, una teoría de los objetos a los que se refieren la cláusulas completivas que complementan verbos que expresan actitudes proposicionales del tipo "X cree que ..." , "X percibe

que ...", "X desea que ...". La investigación de los objetos a los que se refieren las oraciones que complementan esta clase de verbos dará precisamente la clave para dilucidar la estructura lógica de las representaciones mentales. La adscripción de contenido semántico a las representaciones mentales--la instancia intencional--se revela en primer lugar en las oraciones que utilizamos cotidianamente para expresar la atribución de estados psicológicos.

Por otra parte, parece claro que si se necesita elaborar una teoría semántica para los símbolos mentales que se postulan cuando se analiza las entidades sobre las que cuantifica la teoría psicológica, sobre todo teniendo en cuenta que esos símbolos mentales tienen la característica de ser portadores de contenido semántico, lo que resulta fundamental resaltar ahora, al menos según este sumario de ideas, es que parece posible conceder plausibilidad a la tesis de que la teoría semántica para dichos símbolos mentales se puede empezar a construir tentativamente estudiando pormenorizadamente las condiciones de verdad de las adscripciones de actitudes proposicionales que, supuestamente, son su manifestación. Si la adscripción de actitudes proposicionales es esencial para explicar la conducta intencional de los organismos, es importante determinar bajo qué condiciones hacemos estas adscripciones para que ellas sean verdaderas. Es así como la teoría semántica de las actitudes proposicionales nos pone de nuevo sobre el suelo de la concepción de sentido común acerca del ser psíquico. ¿Qué buscamos y qué decimos cuando atribuimos con verdad ciertos estados psicológicos?

Las dos secciones en que está dividido este capítulo corresponden a dos de los pasos fundamentales de la estrategia que ayuda a determinar las propiedades semánticas de las representaciones mentales. La primera parte estará dedicada casi exclusivamente al propósito de seguir con el detenimiento que requiere la excelente teoría que Mark Richard (1990)

ofrece acerca de los múltiples factores que se necesita tener en cuenta si es que se quiere diseñar una teoría suficientemente compleja, rica y plausible acerca de las condiciones de verdad de las adscripciones de actitudes proposicionales. Al hilo de la persecución de este fin, y siguiendo de cerca a este filósofo, se ofrecerán también intuiciones interesantes acerca de cuestiones centrales en la filosofía de la mente contemporánea. Entre los problemas abordados se encuentran, entre otros, el de la naturaleza y estructura de los estados psicológicos que constituyen las actitudes proposicionales mismas, y el de la caracterización de los objetos de creencia.

La segunda parte consistirá en ver cómo los resultados obtenidos en estas investigaciones pueden ser utilizados con provecho para ofrecer una teoría semántica de los símbolos mentales--de las fórmulas del lenguaje del pensamiento--, que vaya un paso más allá de la teoría semántica de las actitudes proposicionales cuando éstas se consideran como parte imprescindible de la psicología de sentido común.

& 2. Decir con verdad lo que creen los demás.

Cuando adscribimos a un sujeto una creencia como, por ejemplo, (a) "Angela cree que Carmen está cansada", consideramos que esta adscripción informa tanto acerca de la *clase* de actitud proposicional en que se encuentra Angela--un estado de creencia en este caso--, como acerca de *qué* es lo que ella cree, a saber: que Carmen está cansada. Inicialmente, un informe verbal como (a) será verdadero si es que su cláusula completiva o enunciado de contenido (content sentence)--en este ejemplo "que Carmen está cansada"-- "representa fielmente (faithfully represents) un enunciado que Angela acepta. Lo que la hace falsa es que no lo represente fielmente" (Richard 1990, 2). De una forma más o menos intuitiva parece que si Angela está en un estado psicológico de creencia tal que dicho estado tiene

un componente o parte que representa a Carmen y otro componente o parte que representa la propiedad de estar cansada, y estas partes están juntas de una manera sintácticamente correcta para que formen el enunciado (mental) de Angela "Carmen está cansada", entonces mi informe verbal "Angela cree que Carmen está cansada" será verdadero si su cláusula completiva representa fielmente el enunciado mental de Angela, esto es, si representa fielmente un enunciado que ella acepta. Precisamente, una de las contribuciones más importantes de la teoría de Richard es la de haber especificado las condiciones fundamentales bajo las cuales el enunciado de contenido de la adscripción de una actitud proposicional representa de una manera fiable el objeto de creencia de Angela o, lo que es lo mismo, un enunciado mental que ella acepta, así como la naturaleza y estructura de ese objeto.

La primera condición necesaria que indica Richard para que haya representación fiel y veraz--quizás la condición más obvia--, "es que mi cláusula completiva y el enunciado mental de <Angela> seleccionen (pick out)", componente a componente, los mismos individuos, propiedades y relaciones..., la misma proposición russelliana", el mismo estado de cosas (1990, 2). En otras palabras, lo que Angela cree y lo que la cláusula completiva nombra tiene que ser la misma proposición russelliana, el mismo estado de cosas. No es desde luego esta afirmación la que presenta más problemas en el conjunto total de la teoría de Richard. Parece claro que mi enunciado "Angela cree que Carmen está cansada" será verdadero cuando su cláusula completiva represente fielmente un enunciado que Angela acepta. A su vez, para que esto suceda, el enunciado de contenido de mi adscripción de esa actitud proposicional a Angela debe nombrar el mismo estado de cosas o proposición russelliana que el estado de cosas que haría verdadero el enunciado que Angela acepta. Si no fuera así, estaríamos hablando de cosas completamente diferentes.

Sin embargo, se podría objetar con razón que esta condición no es, ni con mucho, suficiente. Desde la obra de Frege, las razones son de sobra conocidas por todos. He podido oír decir a Angela que Fernando Pessoa es Fernando Pessoa y saber también que Angela niega el enunciado "Fernando Pessoa es Ricardo Reiss". Sobre estos fundamentos, podría atribuir legítimamente a Angela la creencia "Angela cree que Fernando Pessoa es Fernando Pessoa". Pero si yo dijese ahora que Angela cree que Fernando Pessoa es Ricardo Reiss, diría seguramente algo falso, aunque tanto el enunciado de Angela "Fernando Pessoa es Fernando Pessoa" como el enunciado de contenido de mi adscripción de la última creencia a Angela, esto es, el enunciado "Fernando Pessoa es Ricardo Reiss", nombran o seleccionan la misma proposición russelliana, bajo el supuesto de que el enunciado de identidad es verdadero.

Teniendo muy presente esta clase de argumentaciones, Richard expone la solución que algunos autores han propuesto para evitar los problemas que estos hechos plantean a la semántica de las adscripciones de actitudes proposicionales. Para que haya una relación de representación fiable entre la cláusula completiva de mi adscripción y el enunciado que Angela acepta es preciso tener en cuenta también de qué manera o maneras Angela piensa en Pessoa. Según la terminología que utilizaba en el capítulo precedente, para que haya representación fiable las descripciones que Angela estaría dispuesta a asociar con su uso del nombre "Fernando Pessoa" deben ser las mismas o, en un sentido relevante, prácticamente las mismas, que las descripciones que yo estaría dispuesto a asociar con el uso del mismo nombre. No obstante, Richard encuentra esta solución absolutamente insatisfactoria. En el ejemplo propuesto, cuando atribuyo estados de creencia a Angela, ni busco ni necesito estar seguro de que las maneras en que Angela piensa en Pessoa y las maneras en que yo pienso en Pessoa son las mismas, o, al menos, lo suficientemente semejantes como para que mis

adscripciones sean verdaderas. Dice Richard que la idea misma de "modos de representación" es "demasiado idiosincrática" para que pueda cumplir el papel que se le asigna dentro de la teoría semántica de las adscripciones de actitudes proposicionales (1990, 3). Dos personas, en este caso el adscribidor y el creyente, pueden tener dos vidas cognitivas completamente diferentes, esto es, pueden saber cosas completamente diferentes acerca de Pessoa y, sin embargo, la verdad o falsedad de la adscripción de la creencia "Angela cree que Pessoa es Ricardo Reiss" es, *hasta cierto punto*, independiente de cuáles son las descripciones que el adscribidor y el creyente asocian con el nombre "Fernando Pessoa". La manera propia en que debe interpretarse la reserva "hasta cierto punto" se hará mucho más clara cuando exponga la teoría positiva que Richard ofrece acerca de los objetos de las actitudes proposicionales. De momento, un poco de paciencia.

En cualquier caso, lo que pone ante la discusión crítica esta clase de hechos es que, en numerosos casos, cuando adscribimos con verdad una actitud proposicional, como una creencia, estamos dando también información acerca de *cómo* el creyente cree la correspondiente proposición russelliana y no meramente qué es lo que él cree. Ciertamente, si Angela cree que Twain está muerto--por trabajar con un ejemplo del propio Richard--lo que ella cree es la proposición russelliana:

<Twain, la propiedad de estar muerto>

en donde <Twain> está por el individuo Twain mismo y <la propiedad de estar muerto> está por semejante propiedad, según la notación de las proposiciones que proponen Donnellan y Kaplan. Pero es importante tener también presente que Angela cree esta proposición--el estado de cosas con el que se dice que está relacionada--en virtud de la mediación del enunciado (mental): "Twain está muerto".

Estas primeras caracterizaciones empiezan a arrojar alguna luz sobre el problema planteado. De momento permiten hacer algunas distinciones relevantes, entre ellas las siguientes:

El hacer las actitudes triádicas de esta manera <Angela está relacionada con una *proposición* russelliana por medio de un *enunciado*> nos permite distinguir *lo que* alguien cree, de acuerdo con un russelliano (una proposición russelliana), de *cómo* lo cree (una cuestión de qué mediador(es) ella usa). (Richard 1990, 121).

Después de argumentar en el primer capítulo de su libro que las proposiciones, las entidades que nombran las cláusulas completivas, tienen estructura y extraer la conclusión—a pesar de tener en cuenta algunas precauciones importantes que se analizarán en breve—de que la estructura de las proposiciones refleja, o es isomórfica con, la estructura de los enunciados que las expresan, Richard aborda el problema de buscar los posibles candidatos para ejercer el papel de mediadores de las creencias (1990, 35). Pero antes de desarrollar este punto con más precisión es necesario decir algunas palabras más acerca del tratamiento que reciben en general las proposiciones en la teoría de Mark Richard.

Las proposiciones, tradicionalmente, son entidades cuya utilidad teórica ha estado determinada por cumplir con tres funciones fundamentales (Quine 1974).

(i) A las proposiciones les pertenecen sus condiciones de verdad esencialmente.

(ii) Las proposiciones son el significado de nuestros usos de las oraciones. Cuando se traduce una oración de una idioma a otro, ambas oraciones expresan el mismo significado, la misma proposición. (Que la traducción esté "bien hecha", cualquier cosa que ello signifique, es una condición que, obviamente, se sigue de suyo.)

(iii) Las proposiciones son los objetos de nuestras creencias, deseos, esperanzas, etc., y de las creencias, deseos, esperanzas, etc. de los demás. Son los objetos de nuestras actitudes proposicionales.

Respecto de la consideración de esta tercera función, que parece ser la más asequible de las tres mencionadas y a la que están dedicadas casi en su totalidad las páginas que siguen, Richard piensa que las adscripciones de actitudes proposicionales tienen exactamente los elementos que *prima facie* parecen poseer. Ellas son, efectivamente, enunciados en los que un predicado relacional--creer, desear, decidir, etc.--pone en relación dos términos. En un nivel de generalidad máximo, su forma lingüística es:

a (cree, desea, decide, etc.) que S.

En este sentido, la forma lingüística de una adscripción de una actitud proposicional como, por ejemplo, "Antonio cree que Diseño es un caballo", es sintáctica y semánticamente muy similar a la oración "Antonio abrazó a Manuel". Aquí también un predicado relacional, "abrazar", pone en relación dos términos singulares, "Antonio" y "Manuel". Esta forma paralela de tratar ambas oraciones hace que aquellas que funcionan como complementos directos de los verbos que expresan actitudes proposicionales se puedan caracterizar legítimamente como los *nombres* de ciertas entidades--precisamente de las proposiciones, sea como sea que haya que caracterizarlas--, y que estas entidades que ellas nombran sean típicamente *complejas* (Richard 1990, 5). Contrariamente al nombre propio "Manuel", que aparece en el último ejemplo, la expresión "Diseño es un caballo" tiene un grado de complejidad sintáctica que está completamente ausente en aquél. Los filósofos, sin embargo, se caracterizan por ser sumamente críticos con la apariencia de las cosas y los lingüistas, por su parte, los complementan por ser muy precavidos con la estructura superficial de las oraciones. ¿Por

qué habríamos de sostener confiadamente que las proposiciones tienen una estructura compleja simplemente porque la tengan las expresiones que funcionan como nombres suyos? ¿Qué justifica esta inferencia? Los puntos de vista fundamentales que se pueden defender en relación a esta cuestión son básicamente estos dos:

(i) Las proposiciones, contrariamente a lo que muestran las apariencias, esto es, contrariamente a lo que muestran sus nombres—las cláusulas completivas—no tienen ningún género de estructura interna.

(ii) Las proposiciones sí que tienen estructura, una estructura que precisamente, según expresión semitécnica de Richard, "recapitula" (recapitulates) la estructura de la oración que funciona como cláusula completiva (1990, 8).

El argumento principal que ofrece este autor para concluir que las proposiciones sí que tienen estructura, y que ésta es, básicamente, isomórfica con la de los enunciados de los lenguajes naturales que las expresan y nombran, corre paralelo a la discusión del problema de la deducción. Antes, sin embargo, es preciso introducir una pieza de teoría importante para comprender el alcance de las críticas de Richard y el talante de su propia teoría.

Quienes defienden que las proposiciones no tienen ninguna clase de estructura interna suelen razonar así. Las proposiciones son aquellas circunstancias o mundos posibles que hacen verdadera una oración enunciativa. Las oraciones enunciativas expresan, por su parte, proposiciones. Si las proposiciones no tuvieran ninguna clase de estructura interna, entonces, desde este punto de vista, estos tres enunciados serían nombres de, o expresararían exactamente, la misma proposición: (i) " $p \ \& \ q$ "; (ii) " $\text{no } (\text{no } p \vee \text{no } q)$ "; (iii) " $\text{no } (p \rightarrow \text{no } q)$ ". La razón es clara. Dado que estos tres enunciados son lógicamente equivalentes y, por consiguiente, que no hay circunstancias o mundos posibles en las que uno de ellos sea

verdadero y los otros dos sean falsos, entonces esos enunciados expresan la misma proposición.

Cuando se construyen las proposiciones de esta manera, entonces las proposiciones no pueden tener ninguna estructura que sea similar a la que tienen los enunciados que las nombran. En el ejemplo propuesto, los enunciados " $p \ \& \ q$ ", " $\text{no} \ (\text{no} \ p \vee \text{no} \ q)$ " y " $\text{no} \ (p \rightarrow \text{no} \ q)$ ", son tres nombres para el mismo conjunto de mundos posibles, el conjunto que los hace verdaderos. A este respecto, comenta Richard que la diferencia en estructura o forma lógica de los tres enunciados no puede quedar "reflejada en el conjunto de mundos posibles que los enunciados determinan" (1990, 10). ¿Por qué razón? Se concede que la estructura lógica de esta terna de enunciados es, obviamente, diferente y mutuamente excluyente. Las implicaciones no son ni disyunciones ni conjunciones. Esta estructura lógica, por lo tanto, no puede contribuir de ninguna manera a la individuación de la proposición que constituye las circunstancias en las que esos enunciados son verdaderos, porque si así fuera las diferencias en las estructuras lógicas de los enunciados se deberían reflejar en diferencias en las proposiciones que ellos nombran. Es decir, tres proposiciones diferentes harían verdaderos a esos tres enunciados. Si este último fuera el caso, habría que decir que los enunciados que tienen una forma lógica disyunta nombrarían una proposición diferente que los enunciados que tienen una forma lógica condicional. Dado que es precisamente esta afirmación la que se niega--que las proposiciones tengan una estructura semejante a la de los enunciados que las nombran--, no queda otra posibilidad más que concluir que las proposiciones no quedan en modo alguno individuadas por los rasgos estructurales de los enunciados que las nombran. En otras palabras, que la proposición que esos tres enunciados nombran es tanto la proposición de que $p \ \& \ q$, como la proposición de que $\text{no}(\text{no} \ p \vee \text{no} \ q)$, como la proposición de que $\text{no} \ (p \rightarrow \text{no} \ q)$. Dicha proposición debería

tener, por consiguiente, tanto la forma de una conjunción, como de una disyunción como de una implicación, o, en realidad, ninguna de estas tres formas (Richard 1990, 9-10).

La conclusión de este argumento, que mostraría que las proposiciones son entidades que no están estructuradas al modo en que lo están los enunciados que las nombran, se topa tarde o temprano con el *problema de la deducción*. No es extraño encontrar casos en los que alguien entiende perfectamente lo que dicen las premisas y la conclusión de un argumento lógicamente válido; cree, por otra parte, en la verdad de sus premisas; pero, sin embargo, no cree en la verdad de la conjunción de las premisas con la conclusión. Se suele señalar frecuentemente que muchos estudiantes de los primeros cursos de lógica se encuentran a veces en esta curiosa situación. Un ejemplo aclarará a qué se está aludiendo cuando se habla del problema de la deducción. Uno de los rompecabezas lógicos que más discusión e incredulidad levanta cuando se escucha por primera vez es el siguiente. Un hombre está enseñando unas fotografías a un grupo de amigos y conocidos y, de pronto, alguien le pregunta que de quién es la fotografía que tiene en ese momento en la mano. Su contestación, por lo demás poco habitual y un tanto rebuscada, es ésta: ni hermanos ni hermanas tengo, pero el padre de este hombre—se entiende que del hombre que representa la fotografía—es el hijo de mi padre. ¿De quién es la fotografía? La contestación adecuada a esta pregunta es que el hombre está mirando la fotografía de su propio hijo.

De acuerdo con las consideraciones previas acerca de las proposiciones, los mundos posibles o circunstancias en los que son verdaderas las premisas de este argumento lógicamente válido son exactamente los mismos en los que es verdadera la conjunción de las premisas y de la conclusión. Pero si es verdad que las proposiciones son conjuntos de mundos posibles que no están estructurados como lo están

los enunciados que los nombran y que las cláusulas completivas en las adscripciones de actitudes proposicionales nombran proposiciones así caracterizadas, entonces ha de ser una verdad lógica que quien crea en las premisas de este argumento tiene que creer necesariamente en su conclusión. Sin embargo, esto es lo que con muchísima frecuencia no pasa. Muchas personas que entienden lingüísticamente a la perfección todo lo que se dice en el argumento insisten persistentemente durante un tiempo en no creer en la verdad de la conclusión, aunque crean en la verdad de las premisas.

Según estas consideraciones acerca de la estructura de las proposiciones, parece que hay que distinguir netamente entre qué es aquello que los objetos de nuestras creencias entrañan y qué es lo que entraña creer en lo que creemos. Si Antonio cree que $p \ \& \ q$, entonces, manifiestamente, lo que Antonio cree entraña también que no ($\text{no } p \vee \text{no } q$). Pero la creencia de Antonio no entraña todas las creencias que se derivan a partir del supuesto de que él cree en lo que cree. En este sentido, es completamente diferente hablar de las consecuencias de las creencias que sostenemos que hablar de que las consecuencias de una creencia deben ser ellas también objeto de nuestras creencias. En otras palabras, que las consecuencias de una creencia no tienen por qué ser ellas mismas objetos de creencia (Fodor 1981a, 185). En un principio, no es una tarea en absoluto trivial darse cuenta de todo lo que deductivamente implica creer en lo que creemos.

El argumento general en contra de la afirmación de que las proposiciones no tienen estructura interna le da pie a Richard para asumir precisamente la tesis contraria, a saber, que las proposiciones son entidades estructuradas al modo como lo están los enunciados que las nombran, y no solamente por lo que respecta a su estructura lógica, sino también, como se verá en breve, por lo que se refiere a la presencia de las contrapartes ontológicas de los términos singulares y predicados que

constituyen aquellos enunciados. El avance de alguno de estos resultados posteriores dice que una proposición es una entidad con estructura sintáctica y, posiblemente, formada a partir de los referentes de los términos lingüísticos que aparecen en los enunciados que las nombran.

Sin embargo, antes de llegar a este punto decisivo es importante tomar una precaución más. Si bien es cierto que Richard concluye en general que "ciertos aspectos de la estructura enunciativa se reflejan en la individuación de las proposiciones" conviene, no obstante, tener en cuenta que esta afirmación no implica que debemos terminar necesariamente "con una teoría en la que la estructura de la proposición sea invariablemente isomórfica con la estructura superficial de las cláusulas completivas que la nombran" (Richard 1990, 35). Una oración y un giro idiomático determinado de dicha oración pueden tener distintas estructuras sintácticas. Con todo, en ocasiones resultar útil y necesario transformar una de estas estructuras en la otra y considerarlas finalmente como una única estructura antes de construir la proposición que ellas nombran. Este ejemplo propuesto por el propio Richard se puede afianzar con otro nuevo, si cabe todavía más intuitivo. Para evitar acabar contando con tantas proposiciones como oraciones diferentes puede llegar a generar un determinado lenguaje, nuestras intuiciones lingüísticas nos convencen inmediatamente de que, por ejemplo, las adscripciones (1) "Juan cree que Antonio montó sobre Diseño" y (2) "Juan cree que Diseño fue montado por Antonio" son la misma adscripción. Este hecho se puede explicar fácilmente diciendo que las propiedades lógicas y semánticas de los enunciados de contenido (1') "Antonio montó a Diseño" y (2') "Diseño fue montado por Antonio" nombran la misma proposición y no dos proposiciones diferentes.

El conjunto de las reflexiones precedentes sugiere abiertamente un esbozo de teoría semejante al que a continuación voy a detallar siguiendo de cerca, una vez más, los pasos de Mark Richard. La consideración de las

adscripciones típicas de actitudes proposicionales hace de éstas enunciados en los que un predicado relacional pone en conexión dos términos. Los términos que cumplen la función lingüística de ser los complementos del verbo que expresa la especie de actitud proposicional en que se encuentra el sujeto, es decir, las cláusulas completivas, nombran proposiciones. Estas proposiciones, en un aspecto relevante, están estructuradas, en un doble aspecto, al modo como lo están las cláusulas completivas que las nombran. Por una parte, deben tener ciertos constituyentes que correspondan a los términos y predicados que aparecen en la expresión de las cláusulas completivas. Por otra, deben tener una estructura lógica que, hasta cierto punto, imite a y refleje la de los enunciados que constituyen la cláusula completiva. A esta doble tesis la denomina Richard "sentencialismo semántico" (1990, 38).

Esta primera enunciación de los resultados precedentes invita, a su vez, a considerar y caracterizar de una manera asimismo relacional y compleja los estados psicológicos en los que se dice que está el sujeto u organismo al que se adscribe la actitud proposicional. Estos estados psicológicos son de hecho, en expresión de Richard, los estados que subyacen (underlie) a la adscripción y de cuyo contenido queremos hablar (1990, 39). A estos estados psicológicos y a su contenido se refiere precisamente la adscripción. Una actitud proposicional, como se ha sugerido ya en varias ocasiones, sería, pues, una relación entre un sujeto y una proposición que está estructurada sintácticamente, bajo ciertas reservas, como los enunciados de los lenguajes naturales y, en concreto, como los enunciados que el sujeto utiliza para referirse a esa proposición, y que tiene miembros o partes que corresponden a los términos y predicados que aparecen en dichos enunciados. Esta segunda tesis recibe el nombre de "sentencialismo psicológico" en la teoría de Richard (1990, 38).

El sentencialismo psicológico, como tesis ontológica acerca de la forma de los estados psicológicos mismos, obliga a asignar estructura al estado psicológico que constituye la actitud proposicional. De esta manera, la forma lógica de las oraciones completivas determina el tipo lógico de la actitud proposicional, del estado psicológico. Si las oraciones en función de complemento directo en contextos lingüísticos que expresan actitudes proposicionales tienen una forma lógica atómica, o disyuntiva o conjuntiva, la correspondiente actitud proposicional habrá de tener la misma forma lógica. En este sentido, se puede hablar de creencias o deseos o esperanzas atómicos, disyuntivos, conjuntivos, etc. y, por lo tanto, de grados de complejidad lógica en nuestras actitudes proposicionales.

Después de estos preliminares necesarios en los que se ha tratado acerca de algunas de las propiedades más básicas de las proposiciones en tanto que entidades que nombran las cláusulas completivas, es preciso plantearse de nuevo la pregunta: ¿cuándo será verdad la adscripción de una actitud proposicional? ¿Bajo qué condiciones? Como se dijo al comienzo de este capítulo, una adscripción será verdadera cuando su cláusula completiva represente fielmente un enunciado que acepta el sujeto al que se adscribe la actitud. Como se desprende con cierta facilidad de las consideraciones previas, esta relación de representación fiable depende, por su parte, de una "relación apropiada entre el objeto semántico de creencia que es nombrado por la cláusula completiva y uno de los objetos psicológicos de creencia" del sujeto al que se adscribe la actitud proposicional (Richard 1990, 38). Si el objeto de creencia del sujeto y lo que nombra la cláusula completiva es un estado de cosas—o proposición russelliana—diferente, es claro que la adscripción no puede ser verdadera. Obviamente, la referencia del enunciado de contenido que utiliza el hablante cuando adscribe una actitud proposicional a un sujeto debe coincidir con los valores semánticos, con los referentes, que este sujeto asigna a los términos y predicados del enunciado

que él acepta. Al menos esto ha de ser así si es que el hablante quiere decir algo verdadero acerca de los contenidos de conciencia del otro. (Véase a este respecto también Castañeda 1977, 294).

& 3. Los mediadores de las creencias.

Después de este breve paréntesis--imprescindible para aclarar la noción de proposición que utiliza inicialmente Richard, así como para ver poco a poco su función en la teoría de las actitudes proposicionales--, hay que buscar ahora los candidatos que puedan cumplir adecuadamente el papel de ser los mediadores de las creencias. Esta investigación supone, no obstante, otro paso intermedio previo que es importante analizar. El párrafo anterior estuvo dedicado a la exposición y comentario de por qué y cómo debían estar estructuradas sintácticamente las proposiciones, aquellas entidades con las que estamos relacionados psicológicamente y que atribuimos a los otros cuando decimos que están en una u otra actitud proposicional. Sin llegar a desarrollar este punto con más detenimiento, también se llegó a comentar suficientemente que las proposiciones debían contener ciertos elementos ontológicos que corresponderían a los términos y predicados que aparecen en las cláusulas completivas de las adscripciones de actitudes. Una buena teoría de las proposiciones, en tanto que entidades que nombran las cláusulas completivas y con las que están relacionados psicológicamente los sujetos, debe especificar ante todo cómo los elementos gramaticales más simples que aparecen en el enunciado de contenido contribuyen semánticamente a la proposición que nombra ese enunciado (Richard 1990, 58). Richard propone llamar genéricamente "contenidos" a cualesquiera cosas que sean las que cumplan en la proposición el papel de ser los correlatos o valores de las expresiones simples que aparecen en las cláusulas completivas. En relación a este punto concreto es deseable hacer notar que

la clase de entidades que se pongan en el lugar del contenido hará que la teoría semántica que establece las condiciones bajo las cuales la adscripción de una actitud proposicional es verdadera, varíe correspondientemente y sea más o menos satisfactoria.

¿Qué podrían ser exactamente estos contenidos? Son tres, fundamentalmente, los candidatos que Richard considera más pormenorizadamente. Ninguno de ellos, sin embargo, llega a satisfacerle por completo, porque ninguno de ellos logra enteramente su cometido: ofrecer adecuadamente las condiciones de evaluación semántica de las adscripciones de actitudes proposicionales. Estos tres candidatos son los sentidos fregeanos, los papeles conceptuales y las proposiciones russellianas.

Así, pues, y según esta enumeración, las primeras entidades que acudirían en nuestro auxilio para ser los contenidos de las proposiciones serían los sentidos fregeanos. De acuerdo con las ideas de Frege, las entidades que nombran las cláusulas completivas son *pensamientos*. De esta manera, la adscripción de una actitud proposicional como, por ejemplo, "Bárbara cree que Ana está casada", típicamente será verdadera sólo en el caso de que "la proposición que nombra la cláusula completiva sea idéntica con, o al menos sea muy similar en contenido a, una proposición con la que Bárbara sostenga una relación psicológica (o causal-psicológica) interesante" (Richard 1990, 60).

Richard afirma que el problema de esta estrategia tan utilizada por algunos filósofos es doble. Primeramente, es claro que diferentes personas pueden asociar diferentes modos de representación bien con un nombre propio y, en general, con un término singular, bien con un predicado. Creo que esta tesis, por lo que se refiere sobre todo al caso de los nombres propios, ha quedado suficientemente establecida en los párrafos del capítulo anterior. Como consecuencia de la posibilidad de tener vidas cognitivas dispares, dos sujetos pueden tener diferentes maneras de pensar

en una proposición, si es que es verdad que debemos construir e individualizar las proposiciones a partir de los sentidos fregeanos de los diferentes constituyentes semánticos que forman el enunciado de contenido. Dos personas, por ejemplo quien adscribe la creencia y quien la tiene, pueden tener dos vidas cognitivas completamente diferentes, es decir, pueden saber de Ana como de alguien que posee diferentes propiedades, o pueden asociar diferentes ideas con el predicado "estar casada", y, sin embargo, la verdad o falsedad de la adscripción de la creencia es, hasta cierto punto, independiente de cuál sea el conjunto de descripciones que el adscribidor y el creyente asocian, o estarían dispuestos a asociar, con el nombre propio "Ana" y con el predicado "estar casada", respectivamente.

En segundo lugar, en oposición a la asunción de Frege de que tiene que haber una relación de identidad o de fuerte semejanza entre el contenido-pensamiento del estado psicológico en que se encuentra el creyente y el contenido-pensamiento que nombra la cláusula completiva de quien atribuye la actitud proposicional si es que una adscripción ha de ser verdadera, Richard sostiene que no necesitamos postular "un emparejamiento de contenido cognitivo no referencial (match of nonreferential cognitive content)" entre lo que nombra la cláusula completiva, esto es, el contenido de "que Ana está casada" en el enunciado "Bárbara cree que Ana está casada", y el objeto con el que está relacionado el estado psicológico del creyente" (1990, 60).

Esta afirmación requiere un mínimo de explicación. ¿Qué significa exactamente aquí "emparejamiento de contenido cognitivo no referencial"? Según la teoría de Frege, esto significa básicamente que tiene que haber suficiente *similitud* entre los contenidos cognitivos del creyente y del adscribidor para que la adscripción de la actitud sea verdadera. El adjetivo "no referencial" alude al hecho de que el acuerdo entre los contenidos cognitivos de los dos sujetos no debería establecerse apelando a las

proposiciones russelianas o estados de cosas que nombran los pensamientos fregeanos del creyente y del adscribidor cuando piensan que Ana está casada. Sólo se deberían comparar sus vidas cognitivas y establecer su identidad o semejanza con independencia de los estados de cosas con los que están relacionados psicológicamente, intencionalmente.

Algunas aclaraciones más son aún pertinentes. Richard remarca una y otra vez la idea de que los distintos pensamientos que el creyente y el adscribidor puedan tener acerca de Ana y de la propiedad de estar casada, podrían ser ciertamente relevantes para establecer las condiciones de evaluación semántica de las adscripciones de creencias como "Bárbara cree que Ana está casada". Pero niega rotundamente que una teoría acerca de las condiciones de verdad de las adscripciones de actitudes dependa *enteramente* de la noción de "emparejamiento de contenido cognitivo no referencial".

Si los sentidos son relevantes para la adscripción de actitudes, entonces son relevantes de una manera indirecta y poco obvia. (Richard 1990, 83)

En concreto, rechaza especialmente la idea de que en nuestras prácticas lingüísticas cotidianas busquemos evidencia en favor de la existencia de este emparejamiento o similitud de sentidos fregeanos cuando queremos hablar con verdad acerca de los contenidos de conciencia de los demás. Richard, además, no niega que los sentidos sean entidades objetivas e independientes de la mente que pueden compartir varios sujetos, pero niega que podamos "hacer comparaciones interesantes entre personas acerca de los contenidos cognitivos no referenciales de un enunciado" (1990, 60). Como resultado de todo ello, no se puede conceder que los pensamientos fregeanos sean las entidades que nombran los enunciados de contenido y,

por consiguiente, que sean sólo ellos los objetos de las actitudes proposicionales mismas.

Richard expone todo el problema de esta manera más precisa. Según Frege, las cláusulas completivas en las adscripciones de actitudes proposicionales nombran pensamientos. Brevemente, los pensamientos son modos de pensar acerca de algo y están constituidos a partir de los sentidos de los diversos elementos semánticos que forman las cláusulas completivas. Así, por ejemplo, cuando yo digo "Ana está casada" estoy expresando el pensamiento *que Ana está casada*. A su vez, este pensamiento está formado por el sentido o sentidos que yo asocio con el nombre "Ana", y por el sentido o sentidos que asocio con el predicado "estar casada". Por lo demás, dado que puede haber muchos modos diferentes de pensar en Ana y en la propiedad de estar casada, no puede haber seguridad de que exista un único pensamiento *que Ana está casada* que se exprese siempre cada vez que un hablante utiliza la oración "Ana está casada". Los sentidos varían intersubjetivamente, aunque es posible que los sujetos los puedan compartir en un determinado momento. Como consecuencia de este punto de vista estrictamente fregeano, en una adscripción como (c) "Bárbara cree que Ana está casada", hay que elegir necesariamente entre uno de estos posibles sentidos. En concreto, "una teoría fregeana tiene que hacer algunas elecciones acerca de qué sentido está nombrado por la cláusula completiva en enunciados como (c). ¿Es el sentido que el hablante expresaría con "Ana está casada"? ¿Es el sentido que Bárbara expresaría con ese enunciado, o todavía algún otro sentido?" (Richard 1990, 66).

Los argumentos que expone Richard para ir eliminando cada una de estas tres posibilidades y, a la larga, la teoría fregeana en su conjunto, varían en longitud y complejidad. Los argumentos que siguen a continuación son algo prolijos y requiere un poco de paciencia seguirlos hasta el final. Tienen, sin embargo, el atractivo de su elegancia y rigor.

La primera alternativa disponible dice que la cláusula completiva en la adscripción "Bárbara cree que Ana está casada" nombra el sentido que la persona de cuyo contenido de creencia estamos hablando, en este caso Bárbara, asocia o estaría dispuesta a asociar con el enunciado "Ana está casada". La pega más inmediata y simple que se podría imaginar para eliminar esta opción se puede presentar de esta manera. Ciertamente la adscripción "Nabucodonosor creía que Hesperus no es Phosphorus" es verdadera. Según la teoría que se está discutiendo, la cláusula completiva de esta atribución debería nombrar el sentido que Nabucodonosor asocia con *el enunciado* "Hesperus no es Phosphorus". Pero esto último es falso, dado que Nabucodonosor no asociaba ningún sentido con *este* enunciado, sino, en todo caso, con otro enunciado que sería su contrapartida babilónica (Richard 1990, 67).

Esta obviedad, sin embargo, no parece del todo satisfactoria. Por este motivo Richard procede a sopesar detenidamente si la primera solución para el problema propuesto podría dar adecuadamente las condiciones de evaluación semántica de adscripciones de actitudes encajadas unas en otras o, como se suele decir, múltiplemente embebidas. El argumento elegantísimo de Richard es como sigue. El primer mérito que cualquiera estaría dispuesto a conceder a la teoría semántica de Frege es que es capaz de explicar de una manera sencilla por qué Bárbara puede creer que Hesperus es Hesperus pero no creer que Hesperus sea Phosphorus. Bárbara asocia distintos sentidos con los nombres "Hesperus" y "Phosphorus" y esta diferencia en los sentidos explica por qué Bárbara cree en la verdad del primer enunciado de identidad pero no en la verdad del segundo, aunque ambos nombren el mismo estado de cosas, a saber, la identidad de Venus consigo mismo. La diferencia en sus creencias se explica diciendo que Bárbara no cree que el sentido del primer enunciado sea idéntico con el sentido del segundo enunciado. Supóngase precisamente que es esto lo que Bárbara cree

de hecho y supóngase también que Ana cree, por el contrario, que los enunciados "Hesperus es Hesperus" y "Hesperus es Phosphorus" tienen exactamente el mismo sentido. Supóngase además que este sentido es idéntico con el que Bárbara asocia con el enunciado "Hesperus es Hesperus". Bajo estas suposiciones plausibles, Bárbara puede decir de una forma verdadera:

(1) Yo creo que Hesperus es Hesperus, pero no creo que Hesperus sea Phosphorus.

Ana puede decir con verdad lo que Bárbara cree de esta manera:

(2) Bárbara cree que Hesperus es Hesperus, pero no cree que Hesperus sea Phosphorus.

Antes de seguir adelante con el argumento, es importante hacer notar que en estas dos últimas adscripciones, sus cláusulas completivas deben nombrar ante todo los sentidos que Bárbara asocia con ellas, aunque en el enunciado (2) sea Ana quien hace la adscripción.

Una vez que se ha hecho esta aclaración, podría llegar a suceder en ciertas circunstancias que Ana no sólo esté en condiciones de afirmar con verdad el enunciado (2) –quizás Ana haya oído decir (1) a Bárbara–, sino que sepa realmente que Bárbara cree que Hesperus es Hesperus. Si fuera así, Ana estaría en condiciones de afirmar con verdad el enunciado:

(3) Yo sé que Bárbara cree que Hesperus es Hesperus.

Pero es justamente a partir de aquí donde comienzan los problemas para la teoría de Frege. La primera opción que se está analizando dice que el

enunciado "Hesperus es Hesperus" en (3) nombra el sentido que Ana asocia con ese enunciado. Bajo esta condición inicial, parece claro que (3) se podría reescribir así:

(4) Yo sé que Bárbara cree que Hesperus es Phosphorus.

La sustitución es válida, porque, por hipótesis, Ana asocia efectivamente los mismos sentidos a los enunciados de identidad "Hesperus es Phosphorus" y "Hesperus es Hesperus". Sin embargo, si (4) fuera un enunciado verdadero, entonces Ana podría afirmar a la vez (2) y (4), y, así, sostener el enunciado falso:

(5) Bárbara no cree que Hesperus sea Phosphorus, pero yo sé que Bárbara cree que Hesperus es Phosphorus.

Según este argumento habría que concluir que la primera estrategia ha fallado, porque acaba rindiendo enunciados falsos cuando se analiza la adscripción de actitudes múltiplemente embebidas. Por consiguiente, lo que procede ahora es analizar la segunda alternativa disponible. Las cláusulas completivas nombran el sentido que el sujeto que adscribe el estado psicológico asocia con el enunciado de contenido. Después de exponer los argumentos anteriores es fácil comprender cómo se desarrollaría el argumento para refutar esta segunda posibilidad. No existe ninguna garantía de que cuando yo adscribo a Bárbara la creencia "Bárbara cree que Hesperus es Hesperus" ella asocie los mismos sentidos al enunciado "Hesperus es Hesperus" que asocio yo, aunque la adscripción sea verdadera.

Una vez que se han eliminado las dos primeras opciones, la tercera cae por su propio peso. Esta última dice que las cláusulas completivas no nombran ni el sentido que el adscribidor les asocia ni el que les asociaría

el creyente, sino algún otro sentido. Pero ¿cuál de ellos? Es fácil adivinar qué problemas plantea esta última posibilidad. De cualquier manera que se conteste a esta cuestión, la respuesta parece que habrá de enfrentarse con las mismas objeciones que se presentaron para los dos casos anteriores.

A la luz de estos argumentos, parece que estamos obligados a considerar que los sentidos fregeanos no son las mejores alternativas como candidatos para la mediación de las creencias, dado que las atribuciones de actitudes proposicionales pueden ser verdaderas o falsas sin tener ninguna evidencia especial acerca de que el sentido o sentidos que el creyente asocia con el enunciado *S* es el mismo que, o muy similar a, el sentido o sentidos que el adscribidor asocia con la cláusula completiva "que *S*" en el enunciado "*a* cree que *S*". Una vez más, Richard afirma contra el fregeanismo que no buscamos semejanza o identidad de sentidos en nuestras prácticas lingüísticas en las que adscribimos con verdad actitudes proposicionales a otros sujetos, cosa con la que estoy completamente de acuerdo.

Este filósofo examina ahora otros posibles candidatos que optan a ser mediadores de las creencias. Esta vez es el turno de los roles conceptuales. Para evitar el problema con el que tiene que enfrentarse el fregeanismo en relación con las variaciones de los sentidos que distintos hablantes pueden asociar con una expresión, la teoría semántica puede apelar a la noción de papel conceptual. De acuerdo con este concepto, el papel conceptual de los enunciados de los lenguajes naturales es similar a la noción de sentido de Frege con la ventaja, sin embargo, de no estar tan sujeto a las variaciones de sentido entre los diversos hablantes. De esta manera, en una adscripción de una actitud proposicional como (d): "Odile cree que la nieve es blanca", usamos la cláusula completiva "que la nieve es blanca" para referirnos "indirectamente a un conjunto de enunciados, a aquellos que juegan la misma clase de papel para Odile que "la nieve es

blanca" lo hace para nosotros. El enunciado (d) dice algo así como que Odile cree verdadero uno de aquellos enunciados, en donde aquellos enunciados son aquellos que juegan el papel de la nieve-es-blanca para Odile" (Richard 1990, 80). Sin embargo, Richard también encuentra insatisfactoria esta teoría--inspirada últimamente, como se puede apreciar, en las teorías de Sellars--, porque es muy difícil estar seguros de que el papel conceptual de dos enunciados sea el mismo para dos hablantes, sobre todo si esos enunciados contienen nombres propios (1990, 93). También afirma que, aun cuando se pudiera llevar a cabo esta comparación, parece muy poco prometedor que sólo la identidad o similitud entre los papeles conceptuales de dos enunciados se debiera tener en consideración para establecer las condiciones de verdad de las adscripciones de actitudes proposicionales.

A estas alturas, merece la pena destacar que Richard no niega en modo alguno que un concepto análogo al de sentido fregeano o al de papel conceptual sea en parte relevante para establecer las condiciones de verdad de las adscripciones de actitudes proposicionales. Sólo afirma una y otra vez que en nuestras prácticas lingüísticas cotidianas en las que adscribimos actitudes a un sujeto no hay evidencia de que los hablantes competentes busquen identidad o similitud de sentidos o de papeles conceptuales para adscribir con verdad actitudes proposicionales. En otras palabras, Richard es reacio a aceptar el modo en que se introducen estas nociones cognitivas en la teoría semántica de las adscripciones de actitudes, y su función a la hora de establecer sus condiciones de verdad.

Simplemente no hay un emparejamiento de contenido no referencial entre nuestros enunciados y los estados <psicológicos> de los otros cuando les adscribimos con éxito actitudes. (Richard 1990, 104).

Por consiguiente y después de estas aclaraciones, ¿qué camino queda libre y transitable? Una de las tesis principales de Mark Richard en su excelente libro es que los verbos que expresan actitudes proposicionales son *indexicales*. Que sean indexicales significa, como poco, que su referencia, las entidades que nombran las cláusulas completivas, las proposiciones, varían de contexto en contexto. Para clarificar suficientemente esta tesis es útil trabajar con un tipo de ejemplo cuyas variaciones son ampliamente consideradas por este filósofo.

Imagínese que Odile acepta el enunciado S: "Twain está muerto" y que Mutt sabe que Odile acepta S. Alguien pregunta a Mutt si Odile cree que Twain está muerto. Mutt contesta afirmativamente: "Odile cree que Twain está muerto". Posteriormente y en otro contexto, alguien escucha a Odile proferir el enunciado "me gustaría conocerle", mientras Odile señala a un cuadro que representa a Twain. Jeff, que ha escuchado decir esto a Odile, es abordado y preguntado si Odile cree que Twain está muerto. Dado que Jeff sabe que Odile negaría el enunciado: "él-el hombre que representa el cuadro-está muerto", contesta negativamente: "Odile cree que Twain no está muerto". Respecto de este ejemplo, Richard mantiene que tanto la respuesta de Mutt como la de Jeff son válidas en cada uno de los contextos en que aquellas adscripciones fueron formuladas. Pero si se sostiene esta tesis, parece que Odile puede creer y no creer en el mismo objeto de creencia y tener, por consiguiente, creencias manifiestamente contradictorias. Pero, en realidad, esto último no sucede.

La explicación de Richard de este fenómeno semántico es, brevemente, como sigue. Por una parte, el contexto conversacional en el que tiene lugar la adscripción de Mutt es tal que Mutt estaría contestando a la pregunta: "¿cree Odile que Twain está muerto?". Dadas las intenciones e intereses conversacionales (en un sentido muy amplio) de los hablantes--de Mutt y de quien le pregunta acerca de las creencias de Odile--, Mutt puede

usar la cláusula completiva "que Twain está muerto" como una representación fiable de un enunciado que Odile acepta. Como la adscripción es verdadera en este contexto, el nombre propio "Twain" que aparece en su cláusula completiva ha de contar como una representación aceptable de uno de los mediadores de las creencias de Odile: de aquel que la pone en relación con Twain mismo. En este sentido, el uso del nombre propio "Twain" por parte de Mutt representa el uso del nombre propio "Twain" por parte de Odile (Richard 1990, 134). Por otro lado, la respuesta de Jeff a la cuestión (implícita) "¿se da cuenta Odile de que el hombre que representa el cuadro está muerto?", es también apropiada. Jeff dice que Odile no cree que Twain esté muerto. Dado que Odile rechaza el enunciado: "ese hombre--el hombre que representa el cuadro--está muerto", la cláusula completiva "que Twain no está muerto" en el contexto de Jeff es también una representación fiable y veraz de un enunciado que Odile acepta. En este caso, como la adscripción de Jeff es también verdadera, su uso del nombre propio "Twain" en la cláusula completiva ha de representar el uso de Odile de las expresiones "él" o "ese hombre" en el enunciado "él no está muerto".

Richard señala que la ventaja más inmediata de esta teoría alternativa respecto de las teorías de inspiración fregeana, es que la verdad de una adscripción no está sujeta a las variaciones de sentido que los distintos hablantes pueden asociar con las expresiones que utilizan. Por ejemplo, la única conexión que existe entre la cláusula completiva que usa Jeff para representar el enunciado que acepta Odile y que funciona como mediador de la creencia de Odile-- "él no está muerto" o "ese hombre no está muerto"--, es que tanto el nombre propio "Twain" en la cláusula completiva "que Twain está muerto" como las expresiones "él" o "ese hombre" tienen exactamente el mismo referente. Aparte de esta coincidencia en la interpretación russelliana de esas expresiones no hay por qué buscar

ni postular identidad o fuerte similitud entre los sentidos que Jeff asocia con su uso del nombre "Twain" y los sentidos que Odile asociaría con las expresiones "él" o "ese hombre" para que la adscripción sea verdadera (Richard 1990, 135).

Como ya he indicado, el esquema general que ofrece Richard acerca de las condiciones de evaluación semántica de las adscripciones de actitudes proposicionales supone pensar que los verbos que expresan especies de actitudes, como por ejemplo el verbo "creer", son indexicales. Esto quiere decir en primer lugar que el valor de verdad de la adscripción "Odile cree que Twain está muerto" varía entre los contextos de Mutt y Jeff. En el contexto de Mutt esa adscripción es verdadera. No lo es, en cambio, en el contexto conversacional de Jeff.

No hay cambio de referencia en las expresiones aparte de la referencia del verbo "creer", ni ningún cambio en Odile. Y, en algún sentido importante, "cree" permanece con un significado constante. Si aceptamos todo esto, entonces diremos que "cree" es indexical. (Richard 1990, 107).

Una vez que se ha llegado a este punto, hay dos cuestiones que deben ser contestadas cuanto antes. La primera es, ¿cuáles son entonces las condiciones para que se dé una representación aceptable de un enunciado que Odile acepta, dado que entre estas condiciones no hay que dar cabida exclusivamente a cualquier noción que sea semejante a la de sentido fregeano? La segunda es: ¿deberían ser tenidas en cuenta estas condiciones dentro de una teoría acerca de las condiciones de verdad de las adscripciones de actitudes proposicionales?

En relación a la primera cuestión piénsese en lo siguiente. Es claro que en el contexto conversacional de Mutt, dado que ésta sabía que Odile acepta el enunciado S: "Twain está muerto", y que también conocía las intenciones y expectativas de su interlocutor cuando éste le preguntó si

Odile cree que Twain está muerto, Mutt puede usar la cláusula completiva "que Twain está muerto" para representar S de una manera aceptable. En este contexto no hay ninguna *restricción* que prohíba utilizar el nombre "Twain" que aparece en la cláusula completiva para representar los usos del nombre "Twain" por parte de Odile. Ahora bien, ¿será verdad que la cláusula completiva "que Twain está muerto" representa en el contexto de Jeff un enunciado que Odile también acepta? La respuesta de Richard es que para que la cláusula completiva "que Twain está muerto" represente en este contexto el enunciado de Odile "él está muerto", se requiere "que Odile acepte "estoy señalando a *a* " para algún nombre propio o indexical *a* de Twain" (Richard 1990, 107). Pero, por supuesto, esta condición no se cumple en el contexto de Jeff. Odile no usa el pronombre "él" en el enunciado "él está muerto" como un indexical que se refiera rígidamente a Twain. Sin embargo, aunque esto sea cierto, el contexto conversacional en que tiene lugar la respuesta de Jeff es tal que Jeff puede usar el nombre propio "Twain" para representar los términos singulares que usa Odile y que se refieren rígidamente a Twain, lo sepa Odile o no. La pregunta del interlocutor de Jeff ha sido si Odile se da cuenta de que *Twain* está muerto. Por la forma en que ha sido formulada esta pregunta lo apropiado en este contexto es usar el nombre propio "Twain" como la representación más adecuada del pronombre "él" que aparece en la oración que Odile acepta (Richard 1990, 135). Esto es lo que hace Jeff.

El comentario de Richard a toda esta larga argumentación es muy esclarecedor y sirve como sumario de todas las ideas que han aparecido en ella.

La respuesta adecuada a la cuestión: ¿cómo puede llegar a representar *e* en

Odile cree que ...e...

alguna expresión (o clase de expresión) en los mediadores de las creencias de Odile? parece ser más o menos ésta: Los intereses e intenciones de un hablante (y hasta cierto punto de su audiencia) determinan cómo las expresiones en la cláusula completiva pueden ser y son usadas para representar mediadores de creencias. Algunas veces, por ejemplo, el hablante y la audiencia se fijan en un nombre específico, término, u otra forma de representar algo. Esto es lo que sucede en el caso de Mutt y Jeff. Este interés mutuo produce una restricción en el uso de "Twain" para representar algo acerca del cómo de la creencia de Odile. En el caso de Mutt, una restricción como –Usa "Twain" para representar "Twain"– es operativa. En el caso de Jeff, alguna como –"Twain" tiene que representar sólo términos que Odile asocia con sus presentes experiencias visuales del (cuadro) de Twain–, es operativa. (1990, 135).

Aunque todo esto parezca estar bien argumentado, sin embargo, es importante señalar que la interpretación russelliana de los enunciados "Twain está muerto" y "él está muerto" en el contexto de Jeff es la misma, a saber:

<Twain, estar muerto>

Lo que pone de manifiesto una vez más este hecho es que la mismidad de la interpretación russelliana de dos enunciados no es una condición suficiente, aunque sí necesaria, para establecer las condiciones de verdad de las adscripciones de actitudes proposicionales.

Precisamente, ésta es la tesis a la que Richard se ve compelido cuando investiga el punto de vista de los russellianos acerca de cómo deben ser construidas las proposiciones y cuáles son las condiciones de evaluación semántica de las adscripciones de actitudes proposicionales. Según el punto de vista de los russellianos, tal y como lo presenta Richard, las proposiciones deben construirse *exclusivamente* a partir de los referentes de las partes semánticamente relevantes que forman el enunciado que se refiere precisamente a esa proposición y la nombra. Los russellianos también

defienden que una actitud proposicional como, por ejemplo, una creencia, es una relación diádica entre un sujeto y una proposición russelliana construida como se acaba de mencionar. Desde este punto de vista, una de las tesis fundamentales de los russellianos respecto de las condiciones de verdad de las adscripciones de actitudes es que si las cláusulas completivas de dos actitudes proposicionales nombran la misma proposición russelliana, entonces no puede suceder que una sea verdadera y la otra falsa (Richard 1990, 109-110).

Según se pudo ver en el capítulo anterior cuando se presentó el rompecabezas de Russell, si dos términos singulares tienen el mismo referente, entonces pueden ser sustituidos uno por otro en cualquier enunciado sin que varíe la proposición que determina ese enunciado. Suponiendo que los nombres "Twain" y "Clemens" sean correferenciales, entonces los enunciados "Twain está muerto" y "Clemens está muerto" nombran o determinan la misma proposición:

<Twain, estar muerto>.

Pero entonces, según la segunda tesis russelliana que se acaba de formular, las adscripciones "Odile cree que Twain está muerto" y "Odile cree que Clemens está muerto" serán ambas verdaderas, dado que sus enunciados de contenido nombran la misma proposición russelliana. Lo mismo sucedería con las adscripciones "Bárbara cree que Hesperus es Hesperus" y "Bárbara cree que Hesperus es Phosphorus". Las dos serían verdaderas por la misma razón que en el caso anterior, aunque nuestras intuiciones nos digan abiertamente que la primera puede ser verdadera sin serlo necesariamente la segunda. El problema mayor del russellianismo en este contexto es, precisamente, que no consigue explicar suficientemente estas intuiciones tan inmediatas. En contraposición, la fuerza que tienen las

tesis de Frege es que no parecen traicionarlas. En este sentido, es necesario añadir todavía algo más a la proposición russelliana para obtener la entidad que nombra la cláusula completiva y, por consiguiente, el objeto propio de la actitud proposicional misma.

A pesar de estas objeciones, los russellianos tienen todavía otro argumento más para explicar cómo es posible que los únicos objetos de creencia sean los contenidos o referentes de las expresiones que funcionan como cláusulas completivas y, al mismo tiempo, que las adscripciones "Odile cree que Twain está muerto" y "Odile cree que Clemens está muerto" sean ambas verdaderas. Los russellianos defienden que estos dos enunciados son ambos verdaderos, porque lo que ellos dicen es exactamente lo mismo. Si las intuiciones lingüísticas de los diferentes hablantes sugieren casi constantemente que esos enunciados difieren en su valor de verdad, o pueden hacerlo, los russellianos pueden explicar esta diferencia atribuyéndola al hecho de que ambos tienen *implicaciones pragmáticas* diferentes (Richard 1990, 120). Lo *que* nombran es lo mismo. *Cómo* lo nombran es diferente.

Richard propone un ejemplo muy gráfico para explicar esta diferencia, aunque reconozca que es un tanto tosco. Supongamos que el modo en que tenemos todas nuestras creencias se puede representar mediante una pizarra en la que están escritos los enunciados en cuya verdad creemos. En estas circunstancias es posible imaginar que Odile puede tener escrito en su pizarra (mental) el enunciado "Twain está muerto" pero no tener el enunciado "Clemens está muerto". Con Jane sucede justamente lo contrario. En su pizarra sólo aparece el enunciado "Clemens está muerto" estando el otro completamente ausente. En cualquier caso, es manifiesto que tanto Odile como Jane creen la proposición russelliana:

<Twain, la propiedad de estar muerto>

pero que la creen bajo diferentes enunciados, es decir, usando diferentes mediadores de las creencias (Richard 1990, 121). Una utiliza el enunciado "Twain está muerto" y la otra el enunciado "Clemens está muerto". El comentario de Richard de estas posibilidades es como sigue:

Aunque los russellianos consideran que las actitudes proposicionales son relaciones con proposiciones, admiten que tales relaciones son relaciones mediadas.... La intuición detrás de tales puntos de vista es que uno tiene una proposición en virtud de estar relacionado con (o, en la versión del estado de creencia, estar en) alguna tercera entidad que determina ella misma la proposición. (Hablando estrictamente, es esta tercera entidad, junto con la situación contextual o histórica en que uno se encuentre, la que determina la proposición.) Según esta perspectiva, se puede creer la misma proposición de maneras muy diferentes. Se la puede creer de una manera y no creerla de otra. (1990, 120-121).

De acuerdo con esta consideración, los russellianos estarían en condiciones de diferenciar *qué* es lo que alguien cree de *cómo* lo cree. Lo que se cree es siempre una proposición russelliana. Cómo se cree esa proposición es una cuestión que depende de los mediadores de las creencias, esto es, de los enunciados escritos en la "pizarra mental" por así decir.

El movimiento más característico de los russellianos en relación a las adscripciones de actitudes es afirmar que los mediadores de las creencias—en este tercer caso, los enunciados—y sus posibles implicaciones pragmáticas, bien para el mismo o para diferentes hablantes, no deben ser tenidos en cuenta en una teoría acerca de las condiciones de verdad de estas adscripciones (Richard 1990, 125). Esta afirmación no significa, por supuesto, que los mediadores de las creencias no sean importantes en absoluto. Por el contrario, Richard sostiene muy acertadamente que si las adscripciones de actitudes a los organismos son relevantes para explicar su conducta intencional, entonces "los mediadores de las creencias y los deseos son tan relevantes como las proposiciones que determinan" (Richard 1990, 121).

Por supuesto, los russellianos admiten gustosamente, como se ha visto, que las oraciones que adscriben actitudes a un sujeto no sólo transmiten información acerca de lo que ese sujeto cree, sino también acerca de cómo el sujeto cree en lo que cree. Lo que no están tan dispuestos a admitir es que la información que esa clase de oraciones transmiten acerca del cómo de la creencia sea importante para determinar la verdad o falsedad de la adscripción. Como mucho, un russelliano aceptaría que esa clase de información es relevante para explicar los distintos grados de asentimiento que estamos dispuestos a otorgar a diferentes clases de oraciones. Pero ahí acaba todo. El valor de verdad de las adscripciones sólo depende de la proposición russelliana que nombra el enunciado de contenido. Si dos adscripciones nombran la misma proposición las dos tendrán el mismo valor de verdad.

Sin embargo, Richard sostiene contra los russellianos que la cuestión que se refiere a cómo un creyente está relacionado con el objeto de su creencia sí que es relevante para establecer las condiciones de verdad de una adscripción. No piensa que los mediadores de las creencias—para los russellianos, los enunciados—sean meramente relevantes para el estudio de las implicaciones pragmáticas de las creencias que se tengan. El argumento de más peso para seguir sosteniendo esta tesis es una vez más que las intuiciones de los hablantes competentes sugieren fuertemente que deberíamos considerar diferencias en los mediadores de las creencias como relevantes para establecer las condiciones de verdad de las adscripciones de actitudes. Los ejemplos de los enunciados del tipo "Bárbara cree que Hesperus es Phosphorus" son de mucho peso aquí. Los russellianos no pueden ignorar estos hechos simplemente semi-postulando que lo único que hay en juego en estos casos son las diferentes implicaciones pragmáticas de los enunciados que se aceptan. No obstante y hasta cierto punto, los russellianos puede que tengan razón. No hay por qué aceptar sin más ni

más que la manera en que un sujeto cree en una proposición haya de ser tomada inmediatamente en cuenta a la hora de ofrecer fidedignamente las condiciones de verdad de las adscripciones de actitudes. Aquí no hay por qué legislar nada antes de elaborar la teoría semántica. Pero, como muy bien señala Richard, tampoco se trata de desestimar sistemáticamente la evidencia compartida que aboga en favor de considerar el cómo de una creencia como relevante para establecer aquellas condiciones de verdad.

En todo caso, cuando recapitula las ideas principales del fregeanismo y del russellianismo, Richard extrae la conclusión de que el esquema general de ambas teorías acerca de la adscripción de actitudes es, en el fondo, sorprendentemente similar. Donde los russellianos hablan acerca de las maneras de aprehender una proposición bajo un enunciado del lenguaje natural, los fregeanos prefieren hablar acerca de los modos de representación de, o modos de pensar acerca de, objetos y estados de cosas. Por lo que respecta exclusivamente al problema de las adscripciones de actitudes y a sus condiciones de verdad, la diferencia entre unos y otros, como se ha podido apreciar a lo largo de estas exposiciones previas, es más bien de grado que de fondo. Su discrepancia se centra exclusivamente en saber cuántos elementos, de entre los varios acerca de los cuales informa la adscripción de una actitud proposicional—el cómo y el qué se cree—, deben tenerse realmente en cuenta para establecer sus condiciones de verdad. Los puntos comunes que comparten las teorías fregeanas y russellianas son claros.

Para cada uno de ellos, las creencias y otras actitudes implican una especie de contenido que se individúa en términos referenciales amplios, al igual que algo ampliamente cognitivo en naturaleza, que en algún sentido relaciona al creyente con el contenido referencial de esta actitud. Por lo que se refiere a la semántica de los lenguajes naturales, la diferencia primaria aquí parece venir en respuesta a la cuestión: ¿contribuye la información acerca de los modos de captar una proposición, o modos de pensar en una referencia... a las

condiciones de verdad de las adscripciones de actitudes? (Richard 1990, 124-125).

Tal y como se ha visto, la respuesta de Richard a esta cuestión es afirmativa. Su solución es un compromiso, por lo tanto, entre el fregeanismo y el russellianismo, pero subrayando la idea de que la especificación de los objetos de las actitudes proposicionales ha de ser sensible a las variaciones de contexto. Estas variaciones contextuales son esenciales para una teoría que trate de establecer qué supone que una cláusula completiva sea una representación admisible de un enunciado que el creyente acepta.

La teoría positiva de Richard es que la cláusula completiva "nos da algún referente russelliano--los referentes de sus constituyentes--y algunas expresiones--aquellas <que aparecen> en el enunciado de contenido de la cláusula completiva" (1990, 136). De acuerdo con esta tesis, las entidades que nombran las cláusulas completivas--las proposiciones--son el vástago de dos cosas completamente desemejantes, a saber, algunos elementos lingüísticos y sus interpretaciones russellianas. Richard llama técnicamente a tales cruces "matrices anotadas russellianamente" o "RAMs" (Russellian annotated Matrixes) según la notación que se empleará a partir de ahora para abreviar (1990, 137).

Ahora bien, ¿cómo contribuyen los RAMs a explicar los hechos semánticos con los que se ha trabajado previamente? Para aclarar este punto es necesario pensar de nuevo en el viejo ejemplo de Odile, Mutt y Jeff. Mutt ha oído decir a Odile que Twain está muerto. El interlocutor de Mutt quiere saber si Odile cree que Twain está muerto. Mutt contesta afirmativamente "Odile cree que Twain está muerto". Por su parte, Jeff ha visto cómo Odile señalaba un cuadro que representa a Twain y decía "me gustaría conocerle". Alguien que estaba presente en ese momento pregunta a Jeff si Odile se da cuenta de que Twain está muerto. En este contexto

conversacional y dados los intereses de su interlocutor, Jeff responde correctamente que Odile no cree que Twain esté muerto.

Para Richard, lo que pondría de manifiesto este ejemplo es que cada uno de esos contextos impone ciertas "restricciones" (restrictions) en los diferentes términos que los hablantes pueden utilizar para representar los diferentes mediadores de que se sirve Odile para expresar sus creencias (1990, 135). Tal y como se ha visto, en el caso de Mutt, el interés mutuo de los hablantes—de Mutt y de sus interlocutores—impone una restricción acerca de cómo el término "Twain" puede representar ciertos mediadores de que se sirve Odile para expresar su creencia. La restricción operativa en este caso es: usa el término "Twain" para representar "Twain". En el contexto de Jeff, las cosas cambian muy notablemente. Aquí, el término "Twain" que utiliza Jeff debe representar los términos con los que Odile expresa sus experiencias perceptuales del cuadro de Twain y las creencias que genera esta percepción. La restricción operativa en este caso es: usa el término "Twain" para representar los términos con los que Odile expresa las creencias que genera su experiencia perceptiva.

Sobre la base de estas consideraciones, Richard piensa, siguiendo muy de cerca en este punto a los russellianos, que una cláusula completiva fundamentalmente nos ofrece dos cosas bien diferenciadas. Por una parte, nos da información acerca de cuáles son los referentes russellianos a los que se refieren sus elementos gramaticales. Por otra, nos ofrece un enunciado que representa enunciados que acepta el sujeto al que se adscribe la actitud (Richard 1990, 136-137). Un RAM se contruye a partir de estas dos clases de cosas tan desemejantes.

Así, por ejemplo, el RAM o proposición que nombra el enunciado de contenido "que Twain está muerto" en el contexto de Mutt es:

< < "Twain", Twain > < "está muerto", estar muerto > >

en donde las palabras entre comillas son elementos lingüísticos sintácticamente bien definidos y gramaticalmente independientes en sus funciones, y los otros elementos están respectivamente por sus interpretaciones russellianas, es decir, por su denotación o referente en el caso del nombre propio "Twain" y por la propiedad de estar muerto en el caso del predicado "está muerto".

Una vez que se tiene una idea más precisa acerca de cómo se deben construir los RAMs y antes de ver su utilidad a la hora de establecer las condiciones de evaluación semántica de las adscripciones de actitudes, conviene señalar que los RAMs no son solamente las entidades que nombran los enunciados de contenido, sino los objetos de las actitudes proposicionales mismas. Richard considera que los enunciados que un sujeto admite tienen sus correspondientes interpretaciones russellianas y que, por lo tanto, los diferentes enunciados en los que, por ejemplo, Odile cree se pueden anotar en el modo de los RAMs.

Si hacemos esto con cada uno de los mediadores de su creencia, terminamos con un conjunto de RAMs. Este conjunto codifica todos los factores acerca del creyente que son relevantes para la verdad y falsedad de las adscripciones de creencias a Odile. Llamemos a este conjunto el sistema representacional del creyente o SR. (1990, 137).

Una vez en posesión de estos dos aspectos de la teoría de Richard-- los RAMs son (i) las entidades que nombran las cláusulas completivas y (ii) los objetos de las actitudes proposicionales de los sujetos--, su modo de funcionamiento se concreta de esta manera: dado que Odile acepta el enunciado "Twain está muerto", ella tiene en su sistema representacional el RAM

$f: \langle \langle \text{"Twain"}, \text{Twain} \rangle \langle \text{"está muerto"}, \text{estar muerto} \rangle \rangle$.

Cuando en su contexto conversacional Mutt atribuye a Odile con verdad la creencia: "Odile cree que Twain está muerto" "lo que <está> haciendo es decir (brevemente) que el RAM que determina la cláusula completiva representa uno de los RAMs del creyente" (Richard 1990, 138). Es decir, con la cláusula completiva no sólo está diciendo algo acerca de la proposición russelliana en la que cree Odile, sino también algo acerca de su sistema representacional, algo, en definitiva, acerca de los mediadores de su creencia. La adscripción de Mutt es verdadera porque, dadas las restricciones que impone su contexto de habla, la cláusula completiva "que Twain está muerto" de su adscripción es el nombre de un RAM idéntico con *f*.

¿Qué sucede en el contexto de Jeff? Aquí las cosas no son tan directas. Dado que Odile no acepta el enunciado "él-el hombre que representa el cuadro-está muerto", ella tiene en su SR el RAM

g: <<"no", no> <"él", Twain> <"está muerto", estar muerto>>

Pero en el contexto de Jeff, la cláusula completiva "que Twain no está muerto" en la adscripción verdadera "Odile cree que Twain no está muerto" nombraría el RAM

h: <<"no", no> <"Twain", Twain"> <"está muerto", estar muerto>>

El problema es que el RAM *h* no es idéntico con el RAM *g*. Sin embargo, el RAM *h* que nombra la cláusula completiva de la atribución de Jeff representaría fidedignamente el RAM *g*, porque el nombre "Twain" en *h*, dadas las restricciones operativas que impone este contexto, es una representación fiable del pronombre personal "él" que utiliza Odile. El

contexto en el que Jeff profiere su adscripción de creencia a Odile es tal que impone restricciones en lo que Richard llama "correlaciones" (1990, 138). El contexto conversacional de Jeff impone la correlación:

"Twain"--> "términos que representan las experiencias perceptivas actuales de Odile".

Debido al hecho de que esta restricción es operativa en este contexto, el RAM *b* que nombra el enunciado de contenido "que Twain no está muerto" es una representación fiable del RAM *g* que está en el sistema representacional de Odile. Así se explicaría que el enunciado de contenido de la adscripción "Odile cree que Twain no está muerto" sea una representación fiable y veraz del enunciado que Odile acepta "él no está muerto" en el contexto de Jeff.

En general, "tomada en un contexto--llamémosle *c*--una adscripción de la forma

t cree que *S*

es verdadera si y sólo si el RAM determinado (en *c*) por <el enunciado> *S* representa un RAM en el sistema representacional de lo que *t* nombra (en *c*), bajo una correlación que obedece todas las restricciones operativas en *c*. O, un poco más vagamente, la adscripción es verdadera si y sólo si el RAM *que S* representa uno de los RAMs de *t*, dadas las restricciones del contexto sobre las correlaciones" (Richard 1990, 140).

Esta es, esquemáticamente, la teoría global de Richard. Encuentro, sin embargo, un primer problema en esta teoría, por lo demás tan lúcida, minuciosa y bien construida. Dado que el papel del contexto es tan sumamente importante para establecer qué restricciones son operativas en

cada caso, y dado que Richard no ofrece una teoría detallada acerca de cómo distintos factores contextuales afectan al valor semántico—a los valores de verdad—de las adscripciones de actitudes proposicionales, su teoría necesitaría de algunos complementos importantes en este respecto. Las intenciones comunicativas de los hablantes, lo que ellos saben acerca de la situación de habla (en un sentido muy general), qué es aquello sobre lo que están principalmente interesados cuando hablan, qué es lo que cada uno de ellos sabe acerca de las intenciones del otro y de sus intereses, etc., son factores relevantes para saber qué cláusulas completivas pueden representar fielmente diferentes enunciados que los sujetos aceptan. Richard afirma incluso que los hablantes crean diferentes contextos a medida que la conversación avanza. Este problema, en cualquier caso, no pretende escabullirlo Richard.

Soy consciente de haber ofrecido mucho menos que una teoría completa <sobre este problema>. Ciertamente no he ofrecido una teoría completa acerca de qué clase de intenciones y cosas por el estilo pueden tener esta clase de efecto. Tampoco he intentado decir qué aspectos observables del contexto del oyente se usan para determinar cuáles son las intenciones relevantes del hablante, etc. Esto es algo acerca de lo cual alguien debería dar una teoría. Esto es algo que, en frase agraciada de David Kaplan, se deja como ejercicio para el lector. (1990, 266).

(Sin duda alguna, éste es un ejercicio muy bueno.)

Así, pues, lo que importaría tener es una teoría acerca de los distintos factores contextuales que influyen a la hora de determinar qué cuenta como una representación aceptable de un enunciado que alguien acepta. Esta teoría sería comparable, por ejemplo, a aquellas teorías que tenemos acerca de cómo las variaciones temporales determinan cambios en los enunciados que expresan una cierta proposición. En el caso de los contextos temporales, la proposición que se determina con el uso del

enunciado "hoy está lloviendo" puede expresarse al siguiente día usando el enunciado "ayer estaba lloviendo". Una variación en el contexto temporal está correlacionada con una variación en el tiempo verbal así como con otros cambios gramaticales relevantes--en los adverbios temporales, por ejemplo--si es que se quiere expresar la misma proposición. Una teoría acerca de cómo los contextos (en un sentido amplio) determinan la semántica de las adscripciones de actitudes proposicionales debería parecerse a este esquema de teoría acerca de cómo los contextos temporales determinan los cambios en los enunciados que expresan proposiciones sujetas a esos mismos contextos temporales.

& 4. Aplicaciones de la teoría de los RAMs a algunos problemas semánticos.

Después de formular su teoría acerca de las condiciones de verdad de las adscripciones de actitudes proposicionales, Richard revisa algunas otras cuestiones íntimamente relacionadas con este problema. Entre ellas se encuentran especialmente la de cómo evaluamos múltiples adscripciones de actitudes proposicionales a un mismo sujeto en las explicaciones del comportamiento que siguen el molde típico de la psicología popular "creencia + deseo = acción", y la del análisis de algunos rompecabezas semánticos famosos como los de Pierre-London-Londres y Paderewski-músico-político que propone Kripke (1988).

Para hacer ver el valor de la teoría de los RAMs, me centraré en el análisis que hace Richard del rompecabezas de Kripke que se refiere a las creencias de Peter acerca del hombre de estado polaco, a la vez que excelente músico, Paderewski. Este ejemplo, como el propio Kripke reconoce, puede servir perfectamente bien como representante del problema genérico que se presenta a propósito de la atribución contradictoria, pero

sin fallos lógicos ni errores en los principios semánticos que se utilizan, de creencias a un sujeto (Kripke 1988, 130).

Un buen día Peter oye hablar en un reunión de Paderewski como de un pianista virtuoso y de gran sensibilidad artística. En estas circunstancias de aprendizaje, Peter utiliza referencialmente el nombre "Paderewski" para referirse a Paderewski. Además, Peter cree que está en condiciones de poder identificar, al menos en parte, el referente del nombre mediante las dos descripciones identificadoras con las que ese nombre ha sido introducido en su vocabulario personal, aunque dicho nombre se haya introducido "parásitamente" como se decía en el capítulo anterior. De esta manera, Peter cree en la verdad del enunciado "Paderewski tiene talento musical". Sobre la base de su afirmación de este enunciado, nosotros le atribuimos verdaderamente la creencia:

(1) Peter cree que Paderewski tiene talento musical.

Según la observación que hace Kripke, hay que tener muy en cuenta que en esta adscripción usamos legítimamente el nombre "Paderewski" que aparece en la cláusula completiva para referirnos al músico y hombre de estado polaco Paderewski. Cuando adscribimos esta creencia a Peter somos hablantes que saben cuál es la situación real y que, por tanto, usan el nombre "Paderewski" para referirse al hombre que tiene esas dos propiedades, sabiendo, además, que las tiene.

Algún tiempo después y en otro contexto diferente, Peter aprende que alguien llamado "Paderewski", polaco de nacionalidad, es un político hábil y honrado a carta cabal. Dada la experiencia de Peter con los políticos y con los músicos, no cree que alguien atesore en su persona estas dos excelentes cualidades. Por consiguiente, cuando Peter aprende en este segundo contexto que Paderewski es un buen hombre de estado, estaría

dispuesto a asentir a la verdad del enunciado "Paderewski no tiene talento musical". De la misma manera que antes, nosotros le atribuimos adecuadamente ahora la creencia:

(2) Peter cree que Paderewski no tiene talento musical.

La paradoja que está dispuesto a señalar Kripke es que tanto (1) como (2) son atribuciones verdaderas de cosas en las que Peter cree. Pero, manifiestamente, (1) y (2) atribuyen creencias contradictorias a Peter. La pregunta es: ¿cree Peter, o no cree, que Paderewski tiene talento musical? Este es el rompecabezas del que debería dar cuenta la teoría de Richard. En su explicación la teoría se pone a prueba.

Querer proporcionar una aclaración suficiente de esta paradoja le obliga a Richard a refinar al máximo el modo en que habría que caracterizar y construir los RAMs. La moraleja que extrae este filósofo de semejantes casos paradójicos es que tenemos que construir RAMs usando *muestras de tipos de representaciones mentales en lugar de hacerlo con muestras de tipos de palabras de los lenguajes públicos*. Por ejemplo, el mismo tipo de palabra del lenguaje público "Paderewski" puede ser la expresión de dos muestras de representaciones mentales completamente diferentes. Si utilizara la terminología de que hacía uso en el capítulo anterior, diría que Peter tiene dos representaciones mentales distintas, una determinada por el archivo "Paderewski, el músico" y otra por el archivo "Paderewski, el político".

Según estas distinciones, para que una muestra de un tipo de palabra del lenguaje natural determine la misma representación mental tiene que satisfacer ciertas condiciones "externas" al igual que ciertas condiciones "internas" (Richard 1990, 183). Una muestra mental que funcione como un nombre propio o, en general, como un término singular, cumple la condición "externa" cuando, primero, la palabra que la expresa en el

lenguaje natural se deletrea o pronuncia, al menos dentro de ciertos límites, de la misma manera, y, segundo, cuando pertenece a la misma "cadena causal" de transmisión. Según esta última condición, una muestra de una representación mental satisface la condición "interna" cuando se agrupa con otras representaciones (básicamente predicados o descripciones definidas que funcionan como predicados) que posee ya el individuo que tiene la muestra.

En el caso de Peter, sus muestras del nombre "Paderewski" satisfacen la condición "externa", pero no la condición "interna". Las muestras de Peter de "Paderewski" van a parar al mismo individuo gracias a la posibilidad de desandar varias cadenas causales de transmisión que se originaron en Paderewski. Estas muestras satisfacen, por lo tanto, la condición "exterior". Las dos muestras de las dos representaciones mentales que Peter expresa con el uso del nombre propio "Paderewski" se han originado en Paderewski mismo. También son gramatical y fonéticamente iguales. Pero esas muestras no cumplen la condición "interna", dado que Peter agrupa la representación "es un músico" con parte de sus muestras del nombre "Paderewski" y la representación "es un político" con la otra parte. Peter manipula, por así decir, estas dos muestras como si nombraran objetos diferentes. Como consecuencia, son las muestras de las representaciones mentales las que resultan relevantes y las que habría que tener en cuenta para construir los RAMs. Peter tiene, así, dos RAMs en su sistema representacional. Uno de ellos es:

i: < <"Paderewski-músico", Paderewski> <"está dotado", estar dotado> >

mientras que el otro es:

j: < <"no", no> <"Paderewski-político", Paderewski>
<"está dotado", estar dotado> >

De acuerdo con este esquema general, una adscripción y su negación pueden ser ambas verdaderas *en diferentes contextos*, aunque el término en la posición gramatical de sujeto en la cláusula completiva sea el mismo tipo de palabra en el lenguaje público y cumpla la condición externa, es decir, tenga la misma interpretación russelliana. Así, las adscripciones de las creencias (1) "Peter cree que Paderewski tiene talento musical" y (2) "Peter no cree que Paderewski tenga talento musical", pueden ser ambas verdaderas en diferentes contextos de uso sin que, por otra parte, se cometa ninguna contradicción. Esto se debe a que los RAMs que nombran sus enunciados de contenido cuentan como representaciones válidas de RAMs diferentes en el SR de Peter.

Como ya se ha dicho, los enunciados de contenido de las adscripciones de actitudes verdaderas dicen algo no sólo acerca de la proposición russelliana en la que cree el sujeto, sino también acerca del sistema representacional de ese sujeto. Pero las adscripciones son verdaderas únicamente cuando se respetan ciertas restricciones en las correlaciones. En el caso de Peter hay *dos* correlaciones relevantes que hay que observar. En la adscripción (1) la correlación pertinente es:

"Paderewski" --> "Paderewski, el músico",

mientras que en la atribución (2) la correlación adecuada es:

"Paderewski" --> "Paderewski, el político".

Con estas indicaciones, es fácil entender que el RAM que nombra el enunciado de contenido de la adscripción (1) sea una representación aceptable de un RAM que está en el SR de Peter. El RAM que nombra la cláusula completiva de (1) sería:

k : $\langle \langle \text{"Paderewski"}, \text{Paderewski} \rangle \langle \text{"está dotado"}, \text{estar dotado} \rangle \rangle$.

Ahora bien, aunque en esta cláusula es claro que usamos el nombre "Paderewski" para referirnos al músico y hombre de estado Paderewski y aunque también sea evidente que Peter no usa el mismo nombre con la misma intención, sin embargo, el RAM k que nombra la cláusula completiva de (1) es una representación fiable del RAM i , si es que se respeta la correlación relevante en este contexto. El mismo razonamiento vale para el RAM que nombra el enunciado de contenido de (2).

Peter, por consiguiente, no acepta dos enunciados contradictorios, a pesar de las apariencias, dado que muestras del mismo tipo de palabra en el lenguaje público están por dos clases diferentes de muestras de representaciones mentales, y sólo los tipos de representaciones mentales son relevantes para construir los RAMs.

En cualquier caso, es importante insistir una vez más en que para Richard las atribuciones (1) y (2) son ambas verdaderas sólo relativamente a los contextos de uso. No se trata de que (1) y (2) sean las dos verdaderas sin introducir ninguna matización, sino de que en unos contextos (1) es verdadera y en otros contextos lo es (2). Si no se introdujera esta matización, se estaría sosteniendo que Peter es un ser completamente irracional que no es capaz de darse cuenta de las contradicciones más palpables. Para que Peter fuera un caso de ser irracional en primer grado harían falta dos condiciones. Una de estas condiciones es que aceptara tanto un enunciado de la forma (3) "Paderewski tiene talento musical" como uno de la forma (4) "Paderewski no tiene talento musical". La otra condición, la fundamental, es que considerara que las dos muestras del nombre propio "Paderewski" que aparece en (3) y en (4) seleccionan y se refieren de hecho al mismo individuo Paderewski. Algo parecido a lo que sucedía con Jorge IV y con sus muestras de "Scott" y

de "el autor de *Waverley*". Pero esta última condición es la que no se da en el caso de Peter. Si se diera, entonces las cláusulas completivas de (1) y (2) serían representaciones fiables de (3) y (4), y se podría sostener que Peter acepta realmente enunciados contradictorios. Pero, en este caso, sería necesario sostener que sólo hay *una* correlación válida para construir los RAMs que nombran las cláusulas completivas, a saber: "Paderewski" en la cláusula completiva representa siempre "Paderewski" en el sistema representacional de Peter. Sin embargo, en realidad hay *dos* correlaciones operativas según los contextos. Repito, en un contexto, la correlación operativa es: "Paderewski" en la cláusula completiva de (1) representa fiablemente una muestra de la representación mental "Paderewski, el músico" en el SR de Peter. La otra correlación es: "Paderewski" en la cláusula completiva de (2) representa de una manera válida una muestra de la representación mental "Paderewski, el político" en el SR de Peter.

& 5. El aprendizaje y la retención de las creencias.

En el último capítulo de su libro, Richard trata otros problemas estrechamente relacionados con la lógica y la semántica de las adscripciones de actitudes proposicionales. Entre los problemas que aborda se encuentran, como algunas de las principales contribuciones de su estudio, la explicación de cómo retenemos las creencias a lo largo del tiempo y la aclaración de la manera en que llegamos a tener nuevas creencias verdaderas acerca del mundo gracias a estas creencias que retenemos. También debe destacarse en relación a este asunto, como otra de las aportaciones principales de la teoría de los RAMs, la forma en que la retención de las creencias afecta

al valor de verdad de las adscripciones de actitudes proposicionales a lo largo del tiempo¹.

Pensemos de nuevo en el ejemplo de las creencias de Peter acerca de Paderewski, pero introduciendo en él los problemas de la temporalidad y de la retención de las creencias. Peter cree que Paderewski tiene talento musical y, al mismo tiempo, cree que Paderewski no tiene talento musical. Según las aclaraciones previas, es claro que para Peter el mismo tipo de palabra del lenguaje público, "Paderewski", puede ser la expresión de dos tipos de representaciones mentales diferentes en un tiempo t , a saber, de la representación "Paderewski, el músico" y de la representación "Paderewski, el político". Richard llama a estas dos clases de representaciones, cuando se caracterizan en un tiempo t , "representaciones sincrónicas" de Paderewski. Peter usa en el tiempo t el tipo de palabra pública "Paderewski" para expresar bien muestras de "Paderewski, el músico", o de "Paderewski, el político". Sin embargo, dado que Peter no se da cuenta de que efectivamente hay un único Paderewski en el que se han originado sus dos muestras de "Paderewski", Peter mantiene aisladas dos representaciones sincrónicas de Paderewski en su SR en el tiempo t .

¹ Otro asunto importante que ocupa a Richard es el de la codificación y estructuración de nuestros pensamientos. Aunque no me ocupe de él directamente, por encontrarse un poco al margen del hilo de pensamiento que he estado siguiendo, no quiero, sin embargo, dejarlo pasar completamente por alto. A lo largo de la discusión acerca de si las adscripciones de actitudes violan la ley de Leibniz—argumentos de la forma "si $x=y$, entonces si $\dots x \dots$, entonces $\dots y \dots$ " son válidos—Richard desarrolla un interesantísimo punto de vista acerca de cómo los RAMs codifican información o, más vagamente expresado, cómo nosotros codificamos nuestros pensamientos. Como se ha visto, Richard afirma que los RAMs codifican información no sólo acerca de la proposición russelliana en la que cree un sujeto, sino también acerca de cómo la cree. En numerosos casos, es palpable que la forma concreta en que se consigue hacer esto es reflejando esta información en un "diferencia estructural" de la cláusula completiva. Diferentes estructuras sintácticas codifican diferentes mediadores de las creencias. Por ejemplo, una adscripción de la forma "a cree que $\dots xy \dots$ " no es la misma que una adscripción de la forma "a cree que $\dots yx \dots$ ". Este fenómeno se puede explicar diciendo que en la regimentación de nuestras creencias, las variables no sólo nos proporcionan objetos bajo la extensión de un cuantificador, sino que también determinan enunciados semánticamente diferentes dependiendo del modo en que ponemos juntas las variables.

Convengamos en llamar "RP1" a las muestras de la representación sincrónica "Paderewski, el músico" y "RP2" a las muestras de la representación sincrónica "Paderewski, el político". Bajo estas condiciones iniciales, el aprendizaje ocurre genéricamente cuando dos representaciones o modos de representación que no estaban agrupados juntos en un tiempo t se reconocen explícitamente como dos modos de pensar en el mismo objeto en el tiempo t' , siendo $t < t'$.

Según esta condición necesaria para que se produzca el aprendizaje—para que haya conocimiento en definitiva—, cuando Peter aprende en el tiempo t' que realmente sólo hay un individuo, Paderewski, que reúne en su persona los dones de ser un músico con gran sensibilidad y un político honrado, lo que "hace" es agrupar sus dos muestras de "Paderewski", PR1 y PR2, en una sola. Llamemos "PR3" a la muestra de "Paderewski" en el tiempo t' cuando Peter ya ha aprendido que sólo hay un único Paderewski. Entonces, según esta condición, parece que en el tiempo t' la nueva representación sincrónica PR3 será igual a la suma de PR1 y PR2. De este modo y según la terminología que utiliza Richard, las dos muestras sincrónicas de Paderewski en el tiempo t se convierten en muestras diacrónicas de Paderewski en *el tiempo t'* .

Sin embargo, esta manera de presentar la situación, por motivos que en seguida se harán evidentes, no debe dar inmediatamente pie a considerar que Peter tiene un único modo de pensar en Paderewski cuando ha aprendido en el tiempo t' que el Paderewski del que hablaba cuando decía que era un músico dotado es el mismo Paderewski del que hablaba cuando decía que era un político honrado. La razón principal para hacer esta afirmación es que si se sostuviera esta tesis no habría forma de explicar cómo se producen el aprendizaje y la retención de algunas de sus antiguas creencias.

Imaginemos ahora la situación completa. En el tiempo t , Peter sabe que Paderewski "el músico" dio un concierto magnífico de piano en el año 1889 en un palacio cerca de Viena. Peter asiente, por tanto, al enunciado "Paderewski tenía talento musical en 1889". Nosotros, en el tiempo t , le adscribimos adecuadamente la creencia:

(1) Peter cree que Paderewski tenía talento musical en 1889.

Algún tiempo después, Peter ha aprendido que sólo hay un individuo que se llama "Paderewski" allí donde creía antes que había dos. No obstante, sigue creyendo que este individuo, Paderewski, tenía talento musical en 1889. De esta manera, la anterior atribución de creencia también es verdadera en el tiempo t' .

(1) Peter cree que Paderewski tenía talento musical en 1889.

La cuestión fundamental es saber cuáles son las condiciones de posibilidad para que se produzca la retención de la creencia de Peter acerca de las dotes musicales de Paderewski para que la adscripción (1) sea también verdadera en el tiempo t' . Antes que nada y para que esto suceda, hay que suponer fundamentalmente que Peter sigue manteniendo muestras de PR1 en su SR desde el tiempo t al tiempo t' . Si esto no sucediera, no se podría decir que Peter ha retenido su creencia acerca de Paderewski. Como ya se sabe, las muestras del nombre "Paderewski" en el SR de Peter en el tiempo t son relevantes para construir dos RAMs diferentes, a saber:

d : $\langle \langle \text{"PR1", Paderewski} \rangle \langle \text{"es un músico", ser un músico} \rangle \rangle$

e : $\langle \langle \text{"PR2", Paderewski} \rangle \langle \text{"es un político", ser un político} \rangle \rangle$

La adscripción a Peter en el tiempo t de la creencia de que Paderewski tenía talento musical en 1889 es verdadera, porque Peter tiene el RAM d en su SR.

Ahora bien, en el tiempo t' después del aprendizaje, Peter agrupa, sintetiza o pone juntas PR1 y PR2. Sin embargo, cuando en t' Peter cree también que Paderewski tenía talento musical en 1889, la creencia de Peter en este momento determinaría el RAM:

f : $\langle \langle \text{"PR3", Paderewski} \rangle \langle \text{"es un músico", ser un músico} \rangle \rangle$.

El problema que surge aquí es que aunque haya que afirmar que Peter ha retenido en el tiempo t' su creencia anterior de que Paderewski tenía talento musical en 1889, sin embargo es manifiesto que el RAM f no es el mismo que el RAM d . Por consiguiente, parece que su creencia en t' de que Paderewski tenía talento musical en 1889, es diferente de su creencia en t de que Paderewski tenía talento musical en 1889, y que, por consiguiente, no ha retenido su antigua creencia.

Para evitar esta complicación es necesario tener muy en cuenta cómo se ha construido este último RAM f . En él aparece la notación $\langle \text{"PR3", Paderewski} \rangle$. La muestra sincrónica de la representación mental PR3 de Peter cuando ha aprendido que hay un único Paderewski determina dos representaciones diacrónicas diferentes, a saber PR1 y PR2. Para que Peter retenga en t' su creencia acerca de Paderewski es preciso que PR1 haya estado presente en el sistema representacional de Peter desde el tiempo t . Pero esto es lo que realmente sucede. La muestra de la representación mental sincrónica PR3 está formada por las representaciones diacrónicas PR1 y PR2. Por consiguiente, en el tiempo t' Peter piensa, *en cierto sentido*, dos cosas diferentes acerca de Paderewski (Richard 1990, 230). Puede pensar en él como músico o como político. De esta manera, la aceptación de

Peter en el tiempo t' de una muestra del enunciado "Paderewski tenía talento musical en 1889" permitiría poner dos RAMs en su SR, precisamente los RAMs d y e . En uno de ellos, aparecerá la anotación $\langle \text{"PR1", Paderewski} \rangle$, mientras que en el otro aparecerá la anotación $\langle \text{"PR2", Paderewski} \rangle$. Para que también sea verdadera en el tiempo t' la adscripción "Peter cree que Paderewski tenía talento musical en 1889" y para que esta creencia sea la retención de una creencia anterior, es necesario que uno de estos RAMs esté presente en el SR de Peter desde el tiempo t , mientras que el otro no lo esté (Richard 1990, 230). Pero esto es precisamente lo que sucede. Peter tiene PR1 en su SR desde el comienzo. Por esta razón se puede decir que Peter, después del periodo de aprendizaje, ha retenido su antigua creencia acerca del talento musical de Paderewski.

Hasta ahora todo va bastante bien. Dos muestras de representaciones mentales diferentes determinan dos RAMs diferentes. Peter conserva "huellas" (traces) de ambos RAMs en su sistema representacional. Estas huellas en la memoria permiten que se le pueda atribuir tanto en el tiempo t como en el tiempo t' la creencia acerca de las dotes musicales de Paderewski. La creencia "Paderewski tenía talento musical en 1889" se le puede atribuir a Peter con verdad en los tiempos t y t' , porque Peter tiene huellas de PR1 en su SR entre un tiempo y otro. Con estas indicaciones, Richard puede explicar fácilmente casos semejantes a los de Peter en los que el creyente tiene dos muestras diferentes de representaciones mentales en t , sin saber que son representaciones del mismo individuo, y una muestra mental en t' que determina dos representaciones diacrónicas diferentes, PR1 y PR2. En un caso como éste es fácil describir en qué consiste el aprendizaje (el conocimiento) y es también sencillo saber cómo habría que hablar y representar las creencias de ese individuo antes y después del aprendizaje.

Sin embargo, es importante poner a prueba la teoría de Richard pensando ahora en la situación justamente opuesta. En este caso, Peter tiene también alguna información confusa y errónea en el tiempo t . Su confusión, sin embargo, es la contraria a la que tenía en el ejemplo anterior. En el tiempo t , Peter cree que hay sólo un hombre, Paderewski, que es a la vez músico y hombre de estado. Llamaré otra vez "PR3" a la correspondiente muestra sincrónica en t de la representación mental de Peter "Paderewski, político y músico". Por desgracia, la situación de hecho es que hay dos personas que se llaman "Paderewski". Una de ellas es un músico y la otra es un político. Peter, por consiguiente, tiene alguna información errónea en el tiempo t acerca de esas dos personas. Cree que sólo hay una persona donde en realidad hay dos².

Imaginemos que en esta situación Peter está dispuesto a afirmar el enunciado "Paderewski es un músico". Legítimamente, le adscribimos la creencia:

- (1) Peter cree que Paderewski es un músico.

Más tarde, en un cierto tiempo t' , Peter aprende que efectivamente hay dos Paderewski. Uno de ellos es, por así decir, Paderewski el músico -brevemente Paderewski 1- y el otro Paderewski el político -brevemente Paderewski 2. Llamemos "PR1" a las muestras sincrónicas de Peter de la representación mental "Paderewski el músico" en t' y "PR2" a sus muestras sincrónicas de la representación mental "Paderewski el político" en el mismo tiempo. Sin embargo, después del aprendizaje Peter acepta también

² Como puede apreciarse, este ejemplo es enteramente paralelo al de Patty y Vicky que presentaba en el último capítulo y plantea los mismos problemas que él, nada más que ampliados y aplicados al problema del aprendizaje y de la retención de las creencias.

el enunciado "Paderewski es un músico". Nosotros le adscribimos correspondientemente la creencia:

(2) Peter cree que Paderewski es un músico.

Esta atribución es verdadera, porque el RAM que nombra su enunciado de contenido es una representación fiable de uno de los RAMs de Peter. Después de aprender que hay dos Paderewski, la cláusula completiva de (2) nombra el RAM de Peter:

g: <<"PR1", Paderewski 1> <"es un músico", ser un músico>>.

El problema se plantea cuando, en el tiempo t , la adscripción (2) es también verdadera. Pero ahora necesitamos saber qué RAM tenemos que poner en el SR de Peter en el tiempo t cuando cree que Paderewski es un músico. Algún RAM hay que poner en el SR de Peter, dado que la adscripción "Peter cree que Paderewski es un músico" es verdadera en el tiempo t y, por consiguiente, su cláusula completiva debe ser el nombre de algún RAM que esté alojado en el SR de Peter. Además, para explicar cómo se produce la retención de esta creencia, para explicar que tanto en t como en t' Peter cree que Paderewski es un músico, Peter debe conservar huellas de este RAM, o de alguna de sus partes, en el tiempo t' .

En relación a este problema hay dos posibilidades. La muestra de la representación mental PR3 en t puede ser parte de dos RAMs diferentes.

h : <<"PR3", > <"es un músico", ser un músico>>

i : <<"PR3", > <"es un político", ser un político>>

Dado que la anotación <"PR1", Paderewski 1> que aparece en el RAM *g* hace verdadera a la cláusula completiva del enunciado de creencia (2) en el tiempo *t'* y que Peter retiene esta creencia desde el tiempo *t*, dicha notación tiene que ser una huella bien de la anotación <"PR3", > del RAM *h* o bien de la anotación <PR3, > del RAM *i*. El problema es que ninguna de estas dos anotaciones puede completarse con un referente russelliano, dado que no hay ningún objeto que satisfaga las descripciones que Peter asocia con el nombre "Paderewski" en el tiempo *t*. Es igualmente injustificado rellenar el espacio en blanco que está por la interpretación russelliana de "PR3" tanto con Paderewski el músico--Paderewski 1--como con Paderewski el político--Paderewski 2. De esta manera, parece que no hay ningún elemento en el sistema representacional de Peter que haya sido retenido a lo largo del tiempo y que permita que la adscripción de la creencia (2) sea verdadera tanto en *t* como en *t'*.

Así, pues, tenemos aquí otro caso en el que la teoría de Richard, por lo demás tan excelente y pormenorizadamente argumentada, podría tener algunos problemas al tratar con creyentes que en un cierto tiempo tienen concepciones erróneas, y con las condiciones de verdad de adscripciones de actitudes proposicionales que intentan representar fielmente información falsa.

Además de éste, hay otros problemas con la teoría. El primero de ellos, en cierto sentido una versión distinta del anterior, es que Richard no dice nada concreto acerca de las condiciones de verdad de las atribuciones de creencias cuando una de las partes del enunciado de contenido que expresa la atribución no tiene una interpretación russelliana, es decir cuando en él, por ejemplo, aparecen nombres de personajes ficticios o nombres vacíos, nombres que no tienen referente. Por ejemplo, ¿cuáles son las condiciones de verdad de enunciados como (1): "William cree que Pegaso es blanco"?

De acuerdo con Richard, la adscripción (1) será verdadera cuando el RAM que nombra la cláusula completiva represente fielmente uno de los RAMs en el SR de William. Pero un RAM es una entidad híbrida que se construye a partir de elementos lingüísticos o, mejor dicho, a partir de muestras de representaciones mentales que cumplen las condiciones interna y externa, y de sus contenidos russellianos. De esta manera, el RAM que nombra la cláusula completiva "que Pegaso es blanco" en (1) sería:

j : < <"Pegaso", > <"ser blanco", ser blanco> >

Dado que Pegaso no existe, no podemos rellenar el espacio dejado en blanco en el RAM j con un referente russelliano. Por consiguiente, en este caso sería relevante preguntar si j es realmente un RAM. Pienso que la respuesta adecuada es que, a primera vista al menos, no parece serlo, porque carece de uno de los elementos fundamentales para formar todo RAM, a saber, la interpretación russelliana de uno de los elementos lingüísticos que contribuyen semánticamente a las condiciones de verdad de la atribución de la creencia. Pero si j no es un RAM, entonces la cláusula completiva "que Pegaso es blanco" no puede ser una representación y, mucho menos, una representación fiable de algún RAM en el SR de William. De hecho, si j no es un RAM, no hay ningún RAM en el sistema representacional de William cuando cree que Pegaso es blanco. Pero, ciertamente, William está en un estado psicológico con estructura enunciativa cuando acepta la oración "Pegaso es blanco".

A esta objeción Richard podría contestar simplemente diciendo que ya había advertido desde un principio que no se ocuparía con esta clase de actitudes proposicionales en las que uno de sus elementos no tiene ninguna interpretación russelliana (1990, 4). De todas las maneras, el interés por contestar esta clase de dificultades permanece todavía en pie.

Tal y como he planteado en otra ocasión, una sugerencia inicial que podría satisfacer este interés teórico es que el espacio en blanco en el RAM *j* debería rellenarse con un conjunto de descripciones—"el caballo alado", "el caballo de que habla la mitología", etc.—que podrían considerarse como partes del propio sistema representacional de William, pero partes que deberían considerarse como pertenecientes al mundo. Quizás estas descripciones podrían ser los contenidos de diversos estados mentales respecto de los cuales la tesis general de la superveniencia fuera verdadera. De esta manera, una representación mental podría ser la interpretación russelliana de ciertas expresiones "vacías" que aparecen en la cláusula completiva de la adscripción de una actitud.

Otro problema con la teoría de Richard es el siguiente. Como se ha podido observar, Richard afirma que el enunciado de contenido de las adscripciones de creencias del tipo "Odile cree que Twain no está muerto" nombra RAMs. Estos RAMs son de la forma genérica:

$$<< \text{"no", no} > <a,c> <b,d> >$$

en los que *a* está por un término singular y *b* por un predicado, mientras que *c* y *d* están por sus respectivas interpretaciones russellianas. Está más o menos claro dentro de la teoría cuáles son las interpretaciones russellianas de *a* y de *b*—individuos y propiedades. Pero no sabemos cuáles son las interpretaciones russellianas de las conectivas lógicas—si es que tienen alguna. ¿Qué corresponde a la negación en una proposición russelliana? El problema de la interpretación realista de las conectivas lógicas no es nuevo en las teorías contemporáneas del conocimiento y habría sido muy útil saber qué piensa exactamente Richard acerca de esta cuestión, para que podamos tener un cuadro más detallado de a qué se parecen los RAMs.

Una última preocupación antes de acabar esta parte. La construcción de los RAMs parece que depende de una división "apropiada" de aquellos elementos lingüísticos que he caracterizado al principio de sintácticamente bien definidos y gramaticalmente independientes en sus funciones. Esta forma de dividir los elementos lingüísticos o mentales/lingüísticos que constituyen los RAMs determina los correspondientes elementos de sus interpretaciones russellianas que también constituyen los RAMs. Concedamos que sabemos cómo manejar todas estas divisiones sintácticas para el caso del español y de otras lenguas semejantes al español, esto es, que sabemos cómo construir RAMs al estilo español. Pero no podemos estar seguros acerca de si las estructuras sintácticas de cualquier posible lenguaje natural o de nuestro lenguaje mental (en caso de que exista) permiten una construcción de RAMs en los que la nítida división elemento a elemento entre elementos lingüísticos y sus contrapartes russellianas sea tan clara como lo es en el caso del español.

Apéndice.

& 6. La semántica y la ontología de las actitudes proposicionales.

Como se ha podido comprobar, Richard se mantiene deliberadamente al margen de la cuestión de si su teoría semántica podría tener alguna validez y utilidad especiales para esclarecer los problemas ontológicos que se plantean a propósito de las actitudes proposicionales y de los estados mentales intencionales en general, o, al menos, de los que tienen estructura enunciativa. Lo único que señala es que sus investigaciones podrían encuadrarse muy bien dentro de la psicología popular. Este filósofo piensa que la teoría de los RAMs intenta acomodarse sobre todo al modo en que pensamos normalmente acerca de los contenidos de nuestras actitudes

proposicionales y acerca de los contenidos de las actitudes proposicionales de los demás. Si el problema ontológico entra en algún punto, desde luego lo hace en este nivel descriptivo de la "psicología popular". Pero si fuera esto todo lo que se puede hacer con la teoría semántica de las actitudes proposicionales, entonces sería difícil ver dónde reside el valor de estas investigaciones para los problemas de fundamentación de la psicología como ciencia, y, sobre todo, para el problema de ofrecer una teoría semántica de los símbolos mentales con los que supuestamente están comprometidas ontológicamente las ciencias cognitivas. Como mucho se habría logrado contar con una teoría válida para completar nuestra forma de pensar más inmediata acerca de los objetos de nuestros estados mentales y de los contenidos de conciencia de los demás. Sería relevante para una teoría intersubjetiva del conocimiento. Aquí acabaría todo, aunque sería ya de por sí mucho.

Este último párrafo está dedicado a exponer las tesis de Fodor acerca de las actitudes proposicionales. Fodor, en esta ocasión, aborda el estudio de las actitudes proposicionales con un enfoque y un sentido difícilmente comparable a otros también propuestos por él y de los que se ha hablado en el capítulo segundo. El interés se centra ahora más bien en ver cómo se puede poner en conexión la teoría semántica de las actitudes proposicionales y de las adscripciones de actitudes, con la clase de problemas ontológicos que plantea la psicología cognitiva como ciencia. Sólo así podría entenderse más claramente cómo la investigación de los dos últimos capítulos es realmente importante para los problemas emprendidos al comienzo del libro.

En el ámbito de la filosofía de la mente y de la filosofía del lenguaje, la tesis principal de Fodor en relación a este tema es muy curiosa, por no decir sorprendente. Lo que trata de argumentar este filósofo norteamericano es que los compromisos ontológicos de la psicología

cognitiva a propósito de las actitudes proposicionales se solapan con los resultados de la investigación acerca de las condiciones *a priori* que toda teoría general de las actitudes proposicionales debería cumplir.

Argumentaré que, consideradas juntas, estas condiciones exigen muy a las claras un tratamiento de las actitudes proposicionales como relaciones entre los organismos y las representaciones internas; precisamente el punto de vista al que han llegado independientemente los psicólogos. Por consiguiente, argumentaré que tenemos buenas razones para suscribir la teoría de los psicólogos de una forma independiente, incluso, de las exigencias empíricas que les han llevado a ella. Asumo que esta convergencia entre lo que es plausible *a priori* y lo que se demanda *ex post facto* es en sí misma una razón para creer que la teoría es probablemente verdadera. (Fodor 1981a, 177).

Fodor comienza enumerando cinco puntos básicos que, como condiciones, debería satisfacer toda teoría acerca de las actitudes proposicionales. Como la discusión más pormenorizada de casi todos ellos se hecho ya en las páginas anteriores, me limitaré aquí simplemente a enumerarlos y a decir algunas palabras sobre el único que las reflexiones anteriores no han tocado suficientemente. El interés principal en este momento no es volver de nuevo al centro del debate sobre estas cuestiones, sino ver cómo Fodor las pone en conjunción para tratar de hacer plausible lo que pretende³.

Las cinco condiciones fundamentales son las siguientes.

(i) No hay forma mejor de analizar las actitudes proposicionales que tratándolas como relaciones entre un sujeto y aquello que nombran los enunciados de contenido de las adscripciones de actitudes proposicionales.

³ En la enunciación de estas condiciones me serviré de la terminología que se ha utilizado hasta ahora, para no producir confusiones innecesarias. Nada de lo que se diga depende de este punto. En cualquier caso, cuando haya alguna alteración en este sentido, la indicaré en su lugar correspondiente.

(ii) Las oraciones que expresan adscripciones de actitudes proposicionales guardan un paralelismo en sus propiedades semánticas con las oraciones que expresan lo que, en general, afirma o dice un sujeto. Esta es la condición que llama Fodor "condición de Vendler".

(iii) Una teoría adecuada de las actitudes proposicionales debería dar cuenta de los problemas que surgen a propósito de por qué la sustitución de una oración por otra en la función de cláusula completiva en las adscripciones de actitudes proposicionales, no conserva el valor de verdad de la adscripción, aunque las dos oraciones se refieran al mismo estado de cosas. Fodor denomina "condición de Frege" a esta última condición.

(iv) Los objetos de las actitudes proposicionales tienen estructura sintáctica, forma lógica. "Condición de Aristóteles" bautiza Fodor a esta nueva condición.

(v) Una teoría interesante de las actitudes proposicionales debería poderse relacionar de una manera interesante (mesh) con las teorías empíricas acerca de los procesos mentales (Fodor 1981a, 178-187).

Nadie pone ya en duda—Fodor tampoco lo hace—que las actitudes proposicionales tienen contenido. Por *contenido* de una actitud proposicional hay que entender, como se ha visto, cualquier cosa que sea aquello a lo que se refieren o nombran las oraciones que en los lenguajes naturales funcionan como complementos directos de verbos que expresan actitudes proposicionales. Estos contenidos, por lo demás, han de tener cierta estructura lógica y, por cierto, una estructura que, con ciertas precauciones necesarias, debe reproducir y recapitular la estructura sintáctica de las oraciones de contenido que los expresan. Además, el contenido de una actitud proposicional debe especificarse de tal manera que permita comprender por qué un sujeto puede creer que Hesperus es Hesperus, pero no creer que Hesperus sea Phosphorus, aunque, de hecho, los nombres

"Hesperus" y "Phosphorus" se refieren al mismo objeto: *Venus*. Debe dar cuenta, por consiguiente, de la condición de Frege.

Hasta aquí la teoría de Fodor no se aparta en ningún rasgo esencial de la de Richard. Incluso se puede decir confiadamente que la teoría de Richard está muchísimo mejor elaborada y detallada que la de Fodor. El trabajo de una comunidad de investigadores a lo largo de casi un decenio no es en balde. La teoría de Fodor, sin embargo, se aventura unos pasos más allá con dos reflexiones muy importantes, que son las que deben conducir, en combinación con las demás condiciones mencionadas, a los resultados apetecidos.

La primera de ellas hay que ponerla en relación directa con el modo en que explicamos desde la "psicología popular" la etiología de nuestro comportamiento y la del comportamiento de los demás. Lo que hacemos en esta clase de explicaciones es atribuir a un organismo estados mentales con determinados contenidos y afirmar que el comportamiento, en su más amplia acepción, es el resultado final, entre otras cosas, de las interacciones causales entre estos estados mentales. Por otra parte, estos estados mentales deben estar relacionados no sólo causalmente, sino que deben estarlo, al menos en algunos casos paradigmáticos, por tener el contenido semántico que de hecho tienen. Hay que buscar una teoría de la mente que respete el isomorfismo entre los patrones causales y los patrones semánticos. Desde este punto de vista, puede concederse plausibilidad a la afirmación de que la complejidad estructural de aquello a lo que se refieren las cláusulas completivas debe tener efectos importantes en las relaciones causales que establecen entre sí los estados mentales y, ante todo, las actitudes proposicionales. Este es el primer punto al que aludía antes. Con él la teoría de Fodor se enriquecería y sacaría partido a la teoría semántica general de las actitudes proposicionales.

El segundo punto está contenido realmente en la quinta condición enumerada. La teoría semántica global de las actitudes proposicionales debería ser compatible no solamente con los datos y las intuiciones más inmediatas con las que podemos contar—lingüísticas o no—, sino también con los datos experimentales más refinados que aportan las investigaciones psicológicas. Es decir, la teoría global de las actitudes proposicionales debería ser satisfactoria y coherente no sólo para la psicología popular, sino también para la psicología científica (Fodor 1981a, 186).

Con todos estos elementos a la vista, el optimismo parece ser la mejor actitud respecto del fin deseado.

Estas son, por consiguiente, las condiciones que quiero que cumpla una teoría de las actitudes proposicionales. Argumentaré que, tomadas todas juntas, sugieren fuertemente que las actitudes proposicionales son relaciones entre los organismos y las fórmulas de un lenguaje interno; entre los organismos y <los> enunciados internos. (Fodor 1981a, 187).

La estrategia de Fodor a partir de este momento consiste, primero, en establecer que "los objetos de las actitudes proposicionales son enunciados y, después, <en> mostrar que estos enunciados son probablemente internos" (1981a, 187). Esta terminología, que puede resultar ahora un tanto oscura, se hará más transparente en la discusión que sigue.

En cualquier caso, es conveniente aclarar antes que nada qué quiere decir eso de que los objetos de las actitudes proposicionales son enunciados. En parte, esta labor es algo que ya se ha hecho también en los párrafos anteriores. Si la teoría de los RAMs es adecuada en sus líneas generales, entonces los RAMs, los objetos o contenidos de las actitudes proposicionales que constituyen el sistema representacional de un sujeto y que nombran las cláusulas completivas de las adscripciones de actitudes, contienen necesariamente elementos lingüísticos. Fodor, evidentemente, no

presenta su teoría de esta manera. Su artículo fundamental sobre este asunto tiene casi diez años más que el libro de Richard. Fodor se sirve, sobre todo, de la teoría que Carnap presenta en *Significado y necesidad*, matizándola y corrigiéndola en algunos puntos importantes.

La idea fundamental consiste en mostrar cómo hacer del objeto de una actitud proposicional un enunciado satisface las cinco condiciones antes descritas. (Fodor 1981a, 187-188). De hecho, cuando se consideran así las cosas, todo parece ir bastante bien.

Piénsese en un enunciado de la forma genérica "a cree que S". Si las actitudes proposicionales deben entenderse genéricamente como relaciones, entonces el enunciado "a cree que S" debe interpretarse como una relación entre lo que denota "a" y la oración que funciona como enunciado de contenido, es decir, la oración S. Primer hecho sencillísimo que puede explicar esta teoría. La condición de Vendler también se cumple. Si "a dice que S" y "a cree que S" implican una relación diferente con el mismo enunciado S, entonces cualquiera que sea la estructura lógica y sintáctica que tenga el enunciado S influirá igualmente en las condiciones de verdad de ambas atribuciones. Por otra parte, es fácil imaginar que la condición de Frege también se explica con facilidad. Bárbara puede creer en la verdad del enunciado "Hesperus es Hesperus" y no creer en la verdad del enunciado "Hesperus es Phosphorus". Si los contenidos de las actitudes proposicionales son enunciados, entonces se puede explicar estos hechos acerca de las creencias de Bárbara. Bárbara tiene una relación de creencia con el enunciado "Hesperus es Hesperus", pero no con el enunciado "Hesperus es Phosphorus". Por lo que se refiere a la condición de Aristóteles, ésta se satisface de suyo con esta primera tesis tentativa que dice que los enunciados son los objetos de las actitudes proposicionales. Si las actitudes proposicionales son relaciones con enunciados, entonces la forma lógica del enunciado de contenido determinará la forma lógica y la

complejidad de la actitud proposicional misma. Por fin, si el objeto de una actitud proposicional fuera un enunciado, entonces se podría dar cuenta de algunos hechos empíricos no del todo evidentes que suministra la psicología. Si la complejidad del estado psicológico está en función de la complejidad de aquello que nombran las cláusulas completivas y los procesos mentales son interacciones causales entre estados psicológicos con contenidos complejos, y, por otra parte, algunos comportamientos son el resultado de estas interacciones causales, entonces la teoría psicológica debería ofrecer algunos datos empíricos que confirmaran independientemente que estos comportamientos dependen de la complejidad de los contenidos de los estados psicológicos. En realidad, esto es algo que se puede hacer. Fodor alude en este respecto a los experimentos de Brown y Lenneberg (1954) y Brown (1976). Lo que probarían estos experimentos es que cuanto más complejas sean las descripciones que los sujetos asignan a determinados colores, tanto más difícil les resultará posteriormente identificar de nuevo esos colores en una tarea de reconocimiento. La interpretación de Brown y Lenneberg de estos resultados cuadra perfectamente con lo que habría que esperar de una teoría global de las actitudes proposicionales. El éxito a la hora de reconocer los colores depende de la complejidad de la descripción, y si las descripciones pueden ser los contenidos de ciertos estados psicológicos, entonces los resultados experimentales se pueden explicar a partir de la complejidad de estos estados psicológicos.

Sin embargo, a pesar de todo el buen aspecto que tiene hasta ahora esta teoría, o, más bien, este esquema de teoría, Fodor le encuentra numerosas pegas. Son estas dificultades las que precisamente le conducen a argumentar en favor de la segunda parte de su estrategia. Los contenidos u objetos de las actitudes no son simplemente enunciados, sino *enunciados internos*.

De nuevo hay que decir que algunos de estos obstáculos se han analizado ya en los párrafos anteriores. Siempre se ha indicado que los paralelismos entre el pensamiento y el lenguaje no son perfectos. No siempre que tenemos a la vista dos oraciones con dos estructuras gramaticales diferentes y que funcionan como complementos de verbos que expresan actitudes proposicionales, estamos autorizados a inferir que las adscripciones son diferentes cuando usamos esas oraciones como cláusulas completivas. Es muy plausible pensar que las oraciones "Juan cree que Antonio montó sobre Diseño" y "Juan cree que Diseño fue montado por Antonio" adscriben la misma creencia a Juan, aunque las oraciones en función de complemento no sean del mismo tipo gramatical. Una es activa y la otra es pasiva. Por consiguiente, el criterio para identificar dos adscripciones de actitudes proposicionales no puede ser el mismo que el empleado para identificar dos enunciados (Fodor 1981a, 191).

Otra razón que presenta Fodor, muy sucintamente expresada, para llegar a esta misma conclusión, es la siguiente. Los enunciados son siempre enunciados de un lenguaje natural. Enunciados del español, del inglés, del alemán, del latín o del ruso. Pero parece evidente que alguien puede llegar a creer que Antonio montó a Diseño, aunque no hable ni una palabra de español.

El problema de la falta de isomorfismo entre pensamiento y lenguaje, y el hecho de que se pueda tener la misma creencia aunque las oraciones que complementan un verbo que expresa una actitud proposicional sean oraciones de distintos lenguajes naturales u oraciones de distintos tipos gramaticales--activas, pasivas, etc.--, son indicios que hacen sospechar que los objetos o contenidos de las actitudes proposicionales, para que sigan manteniendo, no obstante, sus propiedades semánticas, han de ser enunciados de un lenguaje que no sea un lenguaje natural. Fodor sostiene,

como es de sobra conocido, que son fórmulas en un Lenguaje Representacional Interno o Lenguaje del Pensamiento (Fodor 1981a, 194).

Una nueva consideración, que tiene que ver sobre todo con la condición (v), apuntala un poco más esta deducción. Esta condición exige que si los enunciados de los lenguajes naturales son los contenidos de las actitudes proposicionales, entonces ha de haber un paralelismo entre la complejidad de la estructura de los enunciados y la complejidad de la estructura de los pensamientos y de los procesos mentales. Esta condición parece cumplirse tal y como lo muestran algunos datos empíricos aportados por la teoría psicológica. Sin embargo, hay otros datos empíricos que desmienten esta conclusión, y Fodor considera que estos hechos son suficientes para cancelar la tesis de que las actitudes proposicionales son relaciones con enunciados de los lenguajes naturales (Fodor 1981a, 193). Fodor recuerda aquí los estudios acerca de las comparaciones transculturales entre los modos de codificar los pensamientos. Según las investigaciones de Brown y Lenneberg a las que se hacía mención anteriormente, la complejidad lingüística de las descripciones que los hablantes ingleses asocian con determinados colores explicaría por qué reconocen posteriormente con mayor o menor facilidad esos mismos colores. Si las descripciones son los objetos de algunas actitudes proposicionales, entonces la complejidad en la descripción explicaría los datos empíricos. Sin embargo, se ha comprobado que la codificación de los pensamientos en inglés *también* predice los patrones de rendimiento de respuesta en las tareas de recuerdo de hablantes que no hablan ni inglés ni ninguna otra lengua de la misma familia. Así sucede con los hablantes del Dani. Dado que los hablantes del Dani no tienen ninguna clase de vocabulario para describir las variaciones cromáticas, no se puede aceptar que las descripciones lingüísticas que los hablantes del inglés asocian con determinados colores sean las mismas que las descripciones de los hablantes del Dani. Lo sorprendente es que, a pesar

de todo, los patrones de rendimiento en las respuestas son prácticamente los mismos. El rompecabezas se plantea cuando se pregunta cómo podrían ser los enunciados del idioma inglés los objetos de las actitudes proposicionales del Dani (Fodor 1981a, 193). Como esta tesis hay que descartarla, dado que los hablantes del Dani no hablan inglés, surge inmediatamente la cuestión de cómo se puede concebir que una propiedad como lo es la de "complejidad en las descripciones lingüísticas" que es necesaria para explicar los datos de los experimentos y que ha sido definida en un principio sólo para los enunciados del inglés, pueda funcionar igualmente bien para el caso del Dani. Respuesta de Fodor: si el sistema representacional que se utiliza para codificar los pensamientos no fuera el lenguaje natural, sino un sistema de representaciones internas que comparten universalmente todos los seres humanos y todos los seres con una psicología semejante, en todo o en parte, a la psicología humana, entonces "no es ya más paradójico, por ejemplo, que la codificación *en inglés* prediga la relativa complejidad de los procesos mentales en Dani" (Fodor 1981a, 197). La culminación de este largo argumento es como sigue.

Por asunción, no es *realmente* la complejidad de los enunciados en inglés lo que predice *nuestro* costo de cómputo (*cost accounting*). No esperaríamos que esa correspondencia fuera algo más que parcial.... Lo que realmente predice nuestro costo de cómputo es la relativa complejidad de las representaciones internas que expresamos cuando usamos oraciones inglesas. Y, de nuevo por asunción, el sistema de representaciones internas subyacente es común tanto a los Dani como a nosotros. (Fodor 1981a, 197).

Así, pues, por un camino o por otro, todos los argumentos de Fodor afluyen en la misma tesis. Los objetos de las actitudes proposicionales son fórmulas del lenguaje interno, del lenguaje del pensamiento. Desde este punto de vista, la intuición que teníamos desde un comienzo de que las adscripciones (1) "Juan cree que Antonio montó

sobre Diseño" y (2) "Juan cree que Diseño fue montado por Antonio" son la misma adscripción, se puede explicar diciendo que las propiedades lógicas y semánticas de los enunciados (1') "Antonio montó a Diseño" y (2') "Diseño fue montado por Antonio" proceden de o dan expresión a la misma fórmula interna que es el objeto propio de la creencia de Juan (Fodor 1981a, 195).

Para aclarar más esta idea, Fodor distingue los siguientes elementos:

(a) Las adscripciones de actitudes como (1) y (2).

(b) Los complementos de esas actitudes como (1') y (2') cuando aparecen en (1) y (2).

(c) Los correspondientes de las adscripciones de actitudes, es decir (1') y (2') cuando aparecen como oraciones propias e independientes, y no en contextos de actitudes proposicionales.

La idea de Fodor es que hay que hacer converger (a), (b) y (c), por un camino o por otro, con la misma fórmula interna, que podemos designar provisionalmente--y de un modo un tanto artificial--, como "B (Antonio montó a Diseño)". Esta expresión hay que interpretarla con el significado de que Antonio está en una relación de creencia con el enunciado interno "Antonio montó a Diseño" (Fodor 1981a, 195). Las oraciones (1') y (2') funcionarían aquí simplemente como *índices* que entresacarían la fórmula "B (Antonio montó a Diseño)" de entre el resto de las fórmulas posibles de ese supuesto lenguaje interno como la fórmula interna con la que Juan tiene una relación de creencia si y sólo si han de ser verdad las adscripciones (1) y (2). En otras palabras, que las oraciones (1') y (2') son las que los hablantes del español utilizan normalmente cuando están en la relación característica que llamamos "creencia" con una muestra del enunciado "Antonio montó a Diseño" y "desean usar una oración del idioma <castellano> para decir qué es lo que ellos creen" (Fodor 1981a, 196).

Las relaciones entre la fórmula interna y las oraciones (1') y (2') se pueden especificar de la siguiente forma.

Las convenciones de los lenguajes naturales funcionan para establecer correspondencias entre sus formas verbales y las fórmulas internas que median las actitudes proposicionales de sus usuarios.... Esta es una forma natural de ver la situación si se piensa en los lenguajes naturales como un sistema de vehículos convencionales para la expresión de los pensamientos. (Fodor 1981a, 196).

Las convenciones del español establecen correspondencias entre "Antonio montó a Diseño" o "Diseño fue montado por Antonio" y la fórmula interna "B (Antonio montó a Diseño)".

En resumen, la propiedades más generales de las actitudes proposicionales parecen demandar entidades del tipo de las sentencias como sus objetos. Estas oraciones no pueden ser oraciones de un lenguaje natural y, por consiguiente, han de ser representaciones internas de un "lenguaje" distinto. Por lo demás, también la construcción de los RAMs sugiere una tesis semejante a ésta. Los RAMs hay que construirlos con muestras de representaciones mentales y no con muestras de tipos de expresiones de los lenguajes naturales.

Ahora bien, si las actitudes proposicionales son relaciones entre los organismos y las fórmulas del Lenguaje del Pensamiento, ¿en qué lugar queda la tesis tradicional de que una actitud proposicional es una relación con una proposición? Es muy interesante constatar que Fodor no piensa que la vieja tesis de que las actitudes proposicionales son relaciones en las que se "captan" o "aprehenden" proposiciones sea una tesis que entre especialmente en conflicto con la suya. En concreto dice:

Estoy tomando seriamente la idea de que el sistema de las representaciones mentales constituye un lenguaje (computacional). *Qua* lenguaje, presumiblemente tiene una sintaxis y una semántica.

Especificar el lenguaje supone decir cuáles son las propiedades en virtud de las cuales sus fórmulas están bien formadas, y qué relación o relaciones existen entre las fórmulas y las cosas en el mundo (no lingüístico). No tengo ni idea de a qué se parecería una semántica adecuada para el sistema de las representaciones internas. Es suficiente que si las proposiciones han de ser tenidas en cuenta en absoluto, lo sean aquí. En particular, nada nos impide especificar una semántica para el SRI <Sistema de Representaciones Internas> diciendo (*inter alia*) que alguna de sus fórmulas expresan proposiciones--esto es, que ellas son relaciones mediadas con proposiciones, y que las representaciones internas sirven de mediadores. (Fodor 1981a, 200).

En cualquier caso, hay que decir que el problema no es tanto la teoría semántica de las representaciones internas, cuanto buscar una teoría semántica *naturalista* de las representaciones mentales. Fodor mismo es consciente de que es un tanto oscuro afirmar que las fórmulas del lenguaje interno "expresan" proposiciones. Sin embargo, él cree que ésta es una dificultad con la que topan las proposiciones, y no las representaciones internas, cuando las proposiciones se contruyen como entidades no estructuradas. Y es un problema de las proposiciones así construidas porque, como vimos anteriormente, si las proposiciones no tuvieran estructura interna o, al menos, no tuvieran la clase de estructura interna que los objetos de las actitudes proposicionales parecen exigir, entonces no se podría dar cuenta de ciertos hechos semánticos.

Las proposiciones neutralizan las diferencias léxico-semánticas entre las distintas maneras de decir la misma cosa. (Fodor 1981a, 201).

Pero el principal obstáculo que Fodor encuentra a la hora de aceptar las proposiciones como los objetos de las actitudes proposicionales es que no ve "cómo un organismo puede estar en una relación (epistémica interesante) excepto si está en una relación (causal/funcional) con alguna

muestra de una fórmula que expresa la proposición" (1981a, 201). La idea fregeana de que las proposiciones se "aprehenden" o "captan" sin decir en qué consiste exactamente esta aprehensión—aparte de la observación de que no es una percepción sensible, porque sus objetos son abstractos, y no es introspección, porque sus objetos no son mentales—, no es avanzar demasiado en la comprensión de nuestras actitudes proposicionales ni en el modo en que nos relacionamos con su objeto cuando estamos en una de ellas. Por ejemplo, con esta doctrina es muy difícil explicar por qué las actitudes proposicionales exhiben relaciones causales.

Por lo que a mí respecta, quiero un *mecanismo* para la relación entre los organismos y las proposiciones, y el único en el que puedo pensar es una mediación con representaciones mentales. (Fodor 1981a, 202).

Es preciso decir que con estas reflexiones se puede empezar a entrever qué clase de relaciones sintácticas tienen las fórmulas del Lenguaje del Pensamiento. La guía, aunque no del todo fiable (como se ha visto), la proporciona la estructura sentencial de las oraciones que funcionan como complementos de las adscripciones de actitudes. Pero el gran caballo de batalla de Fodor desde hace años, como él mismo indica en el texto citado más arriba, es empezar a esbozar una posible teoría semántica que permita la interpretación de estas fórmulas, pero que no traicione la ontología fisicalista que requiere para sí toda ciencia, también la psicológica.

TERCERA PARTE.

EL PROBLEMA DEL SIGNIFICADO.

CAPITULO 5.

TEORIA CAUSAL DE LAS REPRESENTACIONES Y SISTEMAS REPRESENTACIONALES.

& 1. ¿Qué queremos de los estados mentales?

Todo esto está muy bien y da la sensación de que no se ha perdido completamente el tiempo, lo cual siempre es un alivio. El objetivo era, y sigue siendo, la investigación de los estados intencionales. Para aclarar diversos aspectos relevantes de estos estados intencionales fue necesario echar mano de los recursos teóricos que un gran número de disciplinas filosóficas ponen a nuestra disposición. Entender cómo hablamos ingenuamente de algunas de las propiedades de los estados intencionales supuso tan sólo el inicio necesario de un largo recorrido. Después vino una avalancha de reflexiones que quisieron abrir camino y adentrarse en la posibilidad de emprender un estudio científico de estos estados intencionales. En este ámbito había ya un gran número de discusiones heredadas de la tradición en filosofía de la ciencia y en filosofía de la psicología cuyos argumentos fueron utilizados y criticados con el fin de servir a aquella posibilidad. Después se inició una prolija reconstrucción de la teoría semántica, entendida genéricamente como parte integrante de la teoría del conocimiento, que permitiera poner en primer plano cómo han de estar constituidos los estados intencionales cuando se los atribuimos

justificadamente al ser psicológico de un objeto como, por ejemplo, al ser de una persona o de un animal: siempre a cualquier ser que pueda erigirse en el objeto de la investigación psicológica empírica.

De todo este trabajo previo creo que no es exagerado decir que las diversas perspectivas complementarias que han ido apareciendo en los análisis y descripciones han cargado de una manera inevitable la noción de estado mental o de representación mental. Ahora ya no se puede decir alegremente que cada cual rellene estos conceptos de la manera que mejor le parezca. La cuestión, ciertamente, está muy prejuizada en este momento. Así, pues, ¿qué queremos de los estados mentales?

En primer lugar, queremos que sean objetos reales, es decir, queremos que sus propiedades puedan ocupar un lugar propio y coherente dentro de la ontología materialista característica de las ciencias naturales. De los estados psicológicos de los organismos y de las propiedades de estos estados se ocupan la psicología y las ciencias cognitivas como ciencias entre las ciencias especiales. Que sean objetos reales significa, como poco, que pueden ser caracterizados adecuadamente como objetos temporales y espaciales con capacidad causal. En realidad, cuando se habla aquí de esta capacidad causal, se está mencionando un triple aspecto de las características causales de los estados mentales. Los estados mentales son causas y, ante todo, causas de determinados comportamientos. Por su parte, estos estados mentales también pueden ser causados. Finalmente, los estados mentales pueden entrar en relaciones causales, relaciones a las que llamamos genéricamente "procesos mentales". Sin embargo, ésta es sólo la mitad de la historia. La segunda mitad cuenta que los estados mentales son objetos con capacidad representacional. No obstante, esta capacidad representacional sólo se ha tratado hasta ahora desde la perspectiva más inmediata, a saber, atendiendo a las propiedades semánticas de las expresiones que utilizamos para atribuir estados mentales a los organismos. Este es, pienso, uno de los

sentidos fundamentales que hay que dar a la teoría de los RAMs en el conjunto de las investigaciones.

Por desgracia, todas estas consideraciones, como se ha puesto de relieve ya, necesitan de un complemento esencial. Si es verdad que los RAMs son los objetos de las actitudes proposicionales mismas a la vez que los objetos que nombran las cláusulas completivas en las adscripciones típicas de actitudes a los sujetos, y es cierto que los RAMs están constituidos por un elemento lingüístico/mental y por su contraparte russeliana, entonces queda todavía por ver cuál es la relación entre estos dos constituyentes básicos de los RAMs para que su parte lingüístico/mental represente su contraparte russelliana. Esta nueva investigación, sin embargo, debe estar sometida toda ella a una condición fundamental. Las categorías que se utilicen para dar cuenta de esta relación han de ser enteramente compatibles con la ontología realista, materialista y naturalista de las ciencias naturales. Como se suele decir, por pedir que no quede. Sin embargo, el estado de cosas en la investigación de este problema es, quizás, el más controvertido y decepcionante de todos cuantos genera la fundamentación filosófica de las ciencias cognitivas. Lo que presento en los dos próximos capítulos es sólo la discusión y crítica de algunas teorías que, por un camino o por otro, han intentado aportar elementos para solucionarlo. Las condiciones están dadas. Lo que resta es adentrarse con entusiasmo en el nuevo laberinto.

& 2. La teoría semántica naturalista de Dennis Stampe.

El desarrollo reciente de las teorías causales del conocimiento y de las representaciones en general y, concretamente, de las teorías causales de las representaciones lingüísticas y mentales, tiene uno de sus orígenes más

claros y sobresalientes en el famoso artículo de Dennis W. Stampe titulado "Toward a Causal Theory of Linguistic Representations"¹. En este capítulo me propongo presentar alguno de los principales argumentos que llevaron a este filósofo norteamericano a defender una teoría causal de las representaciones en general y una teoría causal de las representaciones lingüísticas en particular. Al mismo tiempo, expondré los problemas más importantes con los que, desde mi punto de vista, tropiezan sus tesis fundamentales, así como las correcciones y distinciones de interés que otros autores han propuesto a sus teorías.

Antes de comenzar, conviene dejar indicado de una forma adecuada que el objetivo último de las teorías causales del conocimiento y de las representaciones es el de proporcionar un marco suficientemente rico donde poder llevar a cabo el programa de la naturalización de las representaciones mentales y lingüísticas y de sus contenidos. Este proyecto hay que inscribirlo dentro de un círculo todavía más amplio de problemas e intereses teóricos, cuyo fin es la construcción de una teoría filosófica naturalista de la mente, es decir, una teoría materialista y realista que sea apta para servir como fundamento de las ciencias cognitivas y, en especial, de la psicología cognitiva. Por "programa de la naturalización de la mente" hay que entender antes que nada aquella investigación que pretende explicar en términos causales los estados mentales o actitudes proposicionales junto con sus contenidos. En otras palabras, el programa de naturalización de los estados mentales trata de explicar la capacidad de representación de estos estados en función de la noción naturalista de causa. Los estados mentales y sus contenidos han de ser objetos del mundo y han de estar sometidos a las mismas leyes causales que el resto de objetos de la naturaleza.

¹ Stampe, Dennis W. 1977. "Toward a Causal Theory of Linguistic Representations." Midwest Studies in Philosophy, Studies in the Philosophy of Language. Eds. Peter A. French et al. Minneapolis: University of Minnesota Press: 42-63.

En la primera parte de este capítulo expondré la teoría naturalista de las representaciones de Dennis Stampe cuya importancia histórica para el desarrollo del problema planteado nunca será suficientemente destacada. Al hilo de esta exposición presentaré alguna de las críticas más sobresalientes que las teorías "ortodoxas" sobre las representaciones han presentado en contra de las teorías causales/naturalistas. Siguiendo a Stampe, argumentaré que el conflicto entre estas dos clases de teorías depende de la importancia relativa que la relación causal entre representación y *representandum* tenga para cada una de ellas. Mientras que para los teóricos naturalistas la representación es, básicamente, un fenómeno causal, los teóricos ortodoxos² defienden que la relación causal no es ni una condición necesaria ni suficiente para que un objeto o estado de cosas sea la representación de otro.

En la segunda parte presentaré el punto de vista de Stampe acerca de aquellos sistemas de representación cuyos elementos tienen "naturalmente" una capacidad intrínseca de representación en virtud de realizar ciertas funciones específicas. Su ejemplo, tan ampliamente discutido, sobre el número de anillos en el tronco de un árbol que representa la edad del árbol, merecerá una consideración especial.

En la última parte, el problema que se planteará será el de las representaciones lingüísticas y su explicación en términos de relaciones causales. En este punto concreto me centraré en el problema de los nombres vacuos (aquellos nombres que designan o mientan un objeto que no existe). La tesis general de Stampe es que esta clase de nombres no se daría en el lenguaje si los mecanismos que regulan su introducción hubieran

² El adjetivo "ortodoxo" puede estar empleado aquí con un matiz al que es difícil no notarle un tono peyorativo que prejuzga un poco la cuestión. Es, sin embargo, el término de que se sirve una y otra vez Stampe para hablar de todas aquellas teorías que no son naturalistas en el sentido mencionado en el texto. Con esta aclaración debería intentarse desterrar aquel posible matiz peyorativo.

funcionado óptimamente. En relación a este problema argumentaré que las críticas de Fodor y Schiffer a la teoría de Stampe son esencialmente adecuadas.

& 3. Las tesis fundamentales.

La teoría general de Stampe acerca de las representaciones se puede resumir en los siguientes puntos fundamentales:

(i) Cuando un objeto R representa a otro objeto O es porque hay una relación causal entre esos dos objetos.

Así, pues, y según este punto (i), el supuesto principal de toda teoría causal de los contenidos de las representaciones es que los miembros de la relación de representación deben ser objetos que puedan entrar en relaciones causales. A primera vista, sin embargo, no en todos y cada uno de los casos en los que parece que estamos autorizados a hablar de representaciones podemos descubrir fácilmente relaciones causales entre la representación y lo representado. Es seguro que tengo la representación de un polígono de 7.521 lados, pero no sé muy bien de qué forma ese objeto ha causado que mi representación tenga las propiedades que de hecho tiene. De todas formas no hay que ser pájaro de mal agüero. Es mejor esperar a ver cómo las teorías causales explican algunos casos de representación más sencillos, porque quizás los casos más complicados se puedan explicar a partir de estos casos más simples. Quizás.

(ii) Normalmente, aunque no necesariamente, en virtud de esta relación causal, cierto conjunto de propiedades F ($f_1...f_n$) del objeto R presenta una relación isomórfica estructural con cierto conjunto de propiedades O ($o_1...o_n$) del objeto O . Un sistema de verdades reguladas contrafactualmente gobierna la producción de esta relación isomórfica de tal forma que, dado que O ha causado R , R no tendría el conjunto de

propiedades F , si O no hubiera tenido el conjunto de propiedades O . Gracias a que R tiene las propiedades F es por lo que (a) R representa a O y (b) podemos aprender algo de O "mirando" a R .

Las propiedades que se expresan en los conjuntos F y O son predicados que se refieren tanto a partes como a relaciones entre estas mismas partes de los objetos R y O (Stampe 1977, 46).

(iii) Como la estructura empírica del mundo se nos aparece como un conjunto enormemente variado de objetos naturales individuales que están relacionados causalmente entre sí de una manera más o menos próxima o remota, es muy plausible que, respecto de esas relaciones causales, se puedan enunciar proposiciones justificadas contrafactualmente del tipo señalado en el párrafo anterior. Por este motivo, es esencial especificar cuál de entre todas estas relaciones causales es la que resulta relevante para que R represente a O . Si no consiguiéramos especificar esta relación causal de una manera plausible, entonces cualquiera de los otros objetos con los que R está relacionado causalmente estaría también representado por él.

Para especificar por qué R representa a O y no a ningún otro objeto que pueda estar también relacionado causalmente con él, es necesario especificar las condiciones de buen funcionamiento o "condiciones de fidelidad" de los mecanismos que están implicados en la producción de las propiedades F de la representación R (Stampe 1977, 49). Si se dan estas condiciones de buen funcionamiento, entonces la representación representaría el objeto que habría causado sus propiedades en aquellas circunstancias.

(iv) Estas condiciones de buen funcionamiento las identifica Stampe explícitamente, al menos para determinados sistemas representacionales, con ciertas condiciones de buen funcionamiento de un mecanismo funcional.

(v) Lo que resulta esencial para las teorías naturalistas de las representaciones es que se puede especificar todas estas condiciones de buen

funcionamiento sin apelar a nociones semánticas o intencionales. Sólo se necesita relaciones naturales, físicas; causales, en una palabra. El contenido de una representación, incluido el de las representaciones lingüísticas, se puede especificar en términos que no son ellos mismos semánticos o intencionales, porque la *especificación del contenido de una representación* consiste precisamente en "un enunciado que dice qué causaría la representación *bajo condiciones relevantes de fidelidad*" (Stampe 1977, 21).

Antes de comenzar con la justificación de estas cinco tesis y de ponerlas a prueba en la explicación de diferentes fenómenos representacionales, es conveniente aclarar la noción general de "sistema representacional" que he introducido en la enunciación de la cuarta tesis. Tengo que advertir antes que Stampe no emplea explícitamente este concepto, aunque nada de lo que se dice a continuación sobre su teoría se ve afectado en lo más mínimo por su utilización. De hecho, la noción de "mecanismo funcional" —de la que sí se sirve explícitamente Stampe en la exposición de su teoría— está íntimamente relacionada con la de "sistema representacional". Concretamente es una de sus especies. Así, pues, el uso de este término está justificado porque con él sería posible referirse a algunos casos de representación que no recoge directamente la teoría de Stampe (bajo el supuesto de que descubriésemos realmente que hay tales casos).

La noción de sistema representacional queda definida de la siguiente manera. Un sistema representacional es "cualquier sistema cuya función es la de indicar cómo un objeto está en relación con algún otro objeto, condición o magnitud. Si, por ejemplo, la función del sistema representacional es la de indicar si *O* está en la condición *A* o *B* y la forma en la que el sistema representacional realiza esta función (*cuando* la realiza) es ocupando uno de dos estados posibles, *a* (cuando indica que *O* es *A*) y *b* (cuando indica que *O* es *B*), entonces *a* y *b* son los elementos

expresivos del sistema representacional y lo que ellos representan (acerca de *O*) es *que* A es (en el caso de *a*) y *que* B es (en el caso de *b*)" (Dreske 1988, 52).

De acuerdo con esta definición, el criterio para clasificar las distintas especies de sistemas representacionales es doble. Por una parte, es preciso saber el modo en el que se establece cuál es la función del sistema representacional. Por otra, hay que determinar cómo los elementos del sistema llevan a cabo esta función.

& 4. Sistemas representacionales del Tipo I.

El primer problema con el que tropieza la teoría de Stampe es que, aparentemente, no todos los sistemas representacionales representan un estado de cosas en virtud de satisfacer todas y cada una de las cinco tesis enunciadas anteriormente. En este sentido, conviene hacer notar que no todos los sistemas de representación cumplen con las condiciones (i) y (ii). Este es el caso de los llamados "signos convencionales". Por ejemplo, si pido a un amigo que me dé indicaciones precisas para ir en moto desde mi casa hasta el teatro donde tengo que comprar entradas para la función de tarde y, en su afán de explicarme cómo llegar hasta allí, utiliza los lapiceros y gomas que tiene sobre la mesa en ese momento para representar con los primeros las calles y con las segundas las plazas por las que tengo que pasar, parece que la función representacional que cumplen los lapiceros y las gomas es meramente temporal y arbitraria. La distribución espacial de los lapiceros y gomas sobre la mesa representa o tiene la función de representar las calles y plazas de mi ciudad de una forma peculiar. Desde luego, los lapiceros y gomas tienen en este ejemplo la función de indicarme las direcciones que tengo que seguir para llegar al teatro, pero también es claro que sólo cumplen con su función en tanto que mi amigo se haya

representado previamente cómo llegar desde mi casa hasta el teatro. Es decir, si mi amigo no tuviera ya antes la información adecuada acerca del camino que tengo que seguir para ir de un lugar a otro y si no tuviera la intención de comunicármela lo más precisamente posible y no lo hiciera así de hecho para que no me pierda, entonces el medio expresivo que utiliza no me informaría por sí mismo acerca del estado de cosas que representa. Este "por sí mismo" significa que los lapiceros y gomas no tienen capacidad o poder *intrínseco* para llevar a cabo la función que mi amigo y yo por convención les hemos asignado temporalmente en nuestro afán de comunicarnos. En otras palabras, estos objetos no tienen "vida representacional" propia, en expresión afortunada de Dreske (1988, 53). Son únicamente instrumentos de representación de los que nos servimos mi amigo y yo. La función de esta clase de sistemas representacionales y la capacidad o poder que tienen sus elementos para realizar esa función *derivan* enteramente de las funciones y capacidades de *otros* sistemas representacionales. En este ejemplo, concretamente, es claro que derivan de dos seres humanos con determinados propósitos comunicativos. Por consiguiente, las relaciones entre las propiedades de los lapiceros y las gomas, y las propiedades de las calles y plazas de mi ciudad—su disposición espacial—, no están reguladas contrafactualmente como de hecho se pide en la tesis (ii). Las propiedades de los lapiceros y gomas, al igual que las relaciones espaciales que esos objetos guardaban antes de servir como instrumentos de comunicación, seguirían siendo las que son, aunque las calles y plazas de la ciudad tuvieran otras propiedades y otras relaciones espaciales diferentes.

A la hora de dar un nombre a los elementos representacionales de esta clase de sistemas, los diferentes autores no parece que hayan llegado a un acuerdo. Hay quienes los llaman simplemente "signos". Otros autores, como Husserl, los llaman "signos indicativos" o "indicaciones" (*Anzeichen*)

sin más (Husserl 1984, 31). Otros los llaman "símbolos" (Dreske 1988, 53). Yo adopto aquí la convención de llamarlos "signos indicativos". En cualquier caso, sea cual sea el nombre que se utilice, lo principal es que un signo indicativo es un objeto o estado de cosas al que se le ha asignado la función de remitir a otro objeto o estado de cosas, porque no tenía la capacidad intrínseca de realizar por sí mismo la función asignada. Esta es la razón de la opinión extendidísima según la cual nada es signo indicativo de suyo, sino que todo signo indicativo lo es porque otros sistemas representacionales le han otorgado esa función.

Ejemplos de signos indicativos son, entre otros muchos, las señales de tráfico, las matrículas de los coches, los códigos en general, los mapas, y también, según Dreske, al menos hasta un cierto punto, los lenguajes naturales. Siguiendo la terminología que utiliza Dreske, llamaré a esta clase de sistemas de representación convencionales, Sistemas del Tipo I (1988, 52).

Una vez hecha esta primera distinción elemental dentro del género sistema representacional, se suscita inmediatamente una serie de preguntas, entre ellas las siguientes: ¿Qué otras clases de sistemas representacionales hay? ¿Hay sistemas representacionales cuyas funciones no les sean asignadas de una forma arbitraria? ¿Hay, por consiguiente, objetos o estados de cosas que tengan la capacidad intrínseca de indicar o remitir a otros objetos o estados de cosas y, por lo tanto, que indiquen de una forma completamente independiente de las funciones que temporalmente les podamos asignar? Y todavía queda una pregunta más, que es completamente decisiva para las teorías causales de las representaciones: ¿se puede explicar de una forma naturalista—es decir, sin apelar a nociones semánticas o intencionales—la forma en la que estos objetos representan? Para responder a estas preguntas con el cuidado necesario es preciso comenzar con el análisis de las cinco tesis establecidas anteriormente en su aplicación al caso de otra clase de sistemas de representación. Me centraré ahora en aquellos

mecanismos funcionales en los que parece que nosotros sólo asignamos *parcialmente* cuáles son sus funciones representativas sin intervenir en absoluto en el modo en que los elementos o estados de esos sistemas representacionales cumplen sus funciones.

& 5. Sistemas representacionales del Tipo II.

La excusa para analizar esta nueva clase de sistemas representacionales y para poner en juego la fuerza explicativa de las cinco tesis anteriores, nos la ofrece el siguiente ejemplo de Stampe (1977, 43). Imagínese un par de gemelos. Uno de ellos, *G1*, hace una fotografía al otro, *G2*. El pensamiento al que se da expresión en la tesis (i) es la suposición de que la fotografía representa a *G2*, porque hay una relación causal entre la fotografía y él. *G2* ha causado de una forma determinada la distribución espacial de los colores que exhibe la fotografía una vez que ha sido revelada. Gracias a esta distribución característica de los colores, la fotografía representa a *G2*. La relación de representación existente entre la fotografía y *G2* es posible porque hay una relación isomórfica estructural entre cierto conjunto de propiedades *O* de *G2* y cierto conjunto de propiedades *F* de la fotografía. La distribución espacial de colores y formas en la fotografía tiene una relación isomórfica estructural con la distribución de colores y formas de *G2*. Esta relación causal entre *G2* y la fotografía está regulada legalmente por los consabidos contrafactuales: la fotografía no habría tenido la distribución de colores y formas que tiene, si no hubiera sido el caso de que *G2* tiene la distribución de colores y formas que de hecho tiene. La tesis (ii) expresa precisamente esta idea.

Indudablemente, aparte de *G2*, otros muchos objetos están relacionados causalmente con la producción de la fotografía, y relacionados, además, de una forma que ejerce la clase de control contrafactual requerido

sobre sus propiedades. *G1* tomó la fotografía. Si él no hubiera apretado el botón de la cámara, entonces la fotografía no habría existido. Si *G1* no hubiera estado un poco bebido después de la fiesta y hubiera enfocado bien, la fotografía no habría tenido las propiedades borrosas que tiene. Este otro conjunto de condiciones circundantes, que podemos identificar quizá con determinadas propiedades o habilidades de *G1*, está también relacionado causalmente con la producción de la fotografía y conectado de una forma relevante-contrafactualmente-con sus propiedades. Stampe reconoce que no solamente *G2*, sino que también *G1* es causalmente responsable de la producción de alguna de las propiedades de la fotografía. Sin embargo, no es en modo alguno deseable acabar con una teoría causal "holista" de las representaciones, según la cual una representación representaría todos y cada uno de los objetos que estuvieran en conexión causal con su producción. Este criterio causal sería un colador demasiado ancho para individuar las representaciones. Seguramente, con él, todo representaría todo de una forma más o menos directa³.

No. La fotografía es la fotografía de un objeto particular. Es la fotografía de *G2* y no de *G1*. Para ilustrar la idea de que no todos los objetos o sucesos que están relacionados causalmente con, y son responsables de, la producción de las propiedades de una representación son representados por ella, Stampe se vale del ejemplo de las percepciones visuales (1977, 43). En la percepción veo un objeto: una manzana, pongamos por caso. Si mi estado visual es parte de una percepción (verdadera), la manzana que yo veo estará allí fuera, en el mundo, y será el objeto que ha causado mi experiencia visual y sus propiedades. Sin embargo, aparte de la manzana, una multiplicidad extraordinaria de objetos diferentes está probablemente relacionada causalmente con la producción de

³ Esta misma queja puede encontrarse en Jaewon Kim, 1977, "Perception and Reference without Causality" (*The Journal of Philosophy* 74: 606-620): 612.

las propiedades de mi experiencia visual. Mis retinas, mis nervios ópticos, miles de neuronas, etcétera. Pero mi estado perceptivo/visual no es una representación de todos estos otros objetos, sino de una manzana. Definitivamente, sería una extravagancia afirmar que mis nervios ópticos están representados en mi experiencia visual. Si ellos estuviesen también representados, habría que decir, como mínimo, que lo estarían de una forma muy inexacta o impropia.

Si esto es así, ¿qué clase de restricciones es preciso imponer a las relaciones causales entre los objetos para que sean unos y no otros los que estén representados por la representación? La respuesta de Stampe, según entiendo, es que hay que poner en conexión las condiciones circundantes que gobiernan la producción de la representación con condiciones de buen funcionamiento de ciertos sistemas funcionales (1977, 51). La fotografía no representa a quien tomó la fotografía, *G1*, porque la determinación de la relación causal relevante entre representación y *representandum* depende de la especificación de las condiciones de buen funcionamiento de la cámara. Y estas especificaciones dependen enteramente de cuál es su función. La cámara tiene la función de hacer ciertas cosas, es decir, ha sido diseñada o pensada para hacer esas y no otras cosas: fotografías. La tesis de Stampe es que la fotografía representa a *G2* y no a *G1*, porque bajo estas condiciones óptimas de funcionamiento, la cámara tiene la función de producir fotografías del objeto que está en relaciones espaciales y causales características con el objetivo, las lentes, la película, y otros elementos de la cámara. Pero sólo *G2* está en estas relaciones espaciales y causales con la cámara, y no *G1*. Bajo condiciones de buen funcionamiento, la cámara "produce representaciones del estado de cosas que se especifica en la cláusula condicional de la especificación de la función del mecanismo" (Stampe 1977, 52). Si la cámara funciona apropiadamente—cumple adecuadamente las funciones para las que ha sido diseñada—, entonces

producirá representaciones (fotográficas) de los objetos que están en relaciones causales específicas con las lentes, la película, etc. La fotografía representa *G2* y no *G1*, porque el conjunto de las propiedades de la fotografía que es relevante para que haya representación ha sido causado precisamente por cierto conjunto de propiedades de *G2* bajo aquellas condiciones de buen funcionamiento de la cámara. Las propiedades de *G1*—estar un poco bebido, ser capaz de tomar fotografías, etc.—no están representadas "propiamente" en la fotografía, porque en la especificación de las funciones de la cámara no se dice nada acerca de que la cámara tenga la función de tomar fotografías de este objeto, aunque él sea causalmente responsable de un determinado conjunto de propiedades de la fotografía.

Hay normalmente un cierto número de cosas representadas en una representación, pero no todas estas cosas serán su objeto "propio" (Stampe 1977, 52).

Según esta terminología, *G1* no es el objeto "propio" que representa la fotografía. Su objeto "propio" es *G2*.

Sin embargo, parece que en estos casos somos nosotros quienes determinamos, al menos hasta cierto punto, para qué es la cámara y, por consiguiente, cuál es el objeto "propiamente" representado en la fotografía de entre todos aquéllos relacionados causalmente con ella. Este es el famoso problema de la teleología de los mecanismos funcionales: el de determinar para qué son y qué representan sus estados funcionales. Este problema, como se comprende fácilmente, está íntimamente relacionado con el de determinar cuál es el objeto propiamente representado en una representación. Ciertamente, las verdaderas dificultades comienzan a partir de aquí.

Stampe hace frente al problema de la teleología de los mecanismos funcionales cuando analiza las diversas funciones que puede llegar a tener

un termómetro. Los termómetros tienen la función--en condiciones de funcionamiento óptimas--de medir la temperatura. En condiciones óptimas, hay una covariación regulada contrafactualmente entre la temperatura de la habitación y la altura de la columna de mercurio en el termómetro. Por este motivo, el estado funcional del termómetro es un indicador fiable de la temperatura del entorno. Sin embargo, podemos usar los termómetros para representar la presión de un volumen constante de aire, en virtud de la legalidad de la relación física entre temperatura, presión y volumen. Con todo, la lectura correcta del estado funcional del mecanismo, del *termómetro*, es "tal o cual temperatura" y no "tal o cual presión atmosférica". Esta situación se hace todavía más patente, piensa Stampe, cuando el mecanismo está integrado en una máquina más compleja. El funcionamiento adecuado de la máquina como un todo depende del funcionamiento adecuado de cada una de sus partes. Más aún, aquello para lo que la máquina es determinará para qué son sus componentes. Si uno de estos componentes es lo que llamamos "termómetro" cuando no está integrado en la máquina, pero su función en la máquina compleja es regular e indicar la presión del aire, entonces el objeto "propio" representado por el estado funcional del mecanismo será la presión atmosférica y no la temperatura. Su lectura correcta será "tal y cual presión" y no "tal y cual temperatura".

Esta teoría sobre el objeto "propio" representado por los estados funcionales de un mecanismo es, pienso, aproximadamente verdadera siempre que escribamos el adjetivo "propio" entre comillas. Es decir, no podemos evitar que la fotografía o el termómetro representen "impropiamente" otros objetos distintos de aquel que es su objeto "propio". La fotografía puede representar impropiamente las propiedades de *G1*, al igual que el termómetro puede representar impropiamente la presión del aire. Y es posible que estos objetos impropiamente representados se conviertan en los objetos propiamente representados por el estado funcional del mecanismo

si las circunstancias o nuestros intereses y propósitos fuesen diferentes. Dado que cuando usamos un termómetro como un termómetro no estamos primariamente interesados--excepto, quizá, de una forma colateral o derivada--en otras condiciones del ambiente, somos nosotros quienes asignamos al mecanismo la función de indicar la temperatura de la habitación y, por consiguiente, de establecer que el estado funcional del mecanismo proporciona determinada información: tal y cual temperatura. De esta manera, ignoramos voluntariamente cualquier otra información adyacente que el mecanismo proporcione necesariamente en virtud de su constitución física, de lo que puede hacer y de las leyes físicas que gobiernan todas las magnitudes relacionadas. Si aceptamos que los mecanismos llamados genéricamente "instrumentos" representan aquel estado de cosas que tienen la función de indicar, y aceptamos también que, a la luz del ejemplo del termómetro, somos nosotros los que determinamos hasta cierto punto estas funciones, entonces parece que somos nosotros los que establecemos lo que el instrumento representa.

Merece la pena subrayar que esta determinación de las propiedades representacionales de un mecanismo funcional en virtud de nuestros intereses está confinada, sin embargo, por la clase de cosas que el mecanismo puede representar en razón de su constitución física y de sus estados funcionales. Es porque el termómetro tiene la constitución que tiene por lo que podemos usarlo como un mecanismo para medir la temperatura o la presión del aire pero no, por ejemplo, para cumplir las funciones que tendría un galvanómetro.

Es por esta razón por la que lo que el indicador (*gauge*) representa es *en parte* convencional, es *en parte* cuestión de lo que decimos que representa.... Sin embargo, esta dependencia respecto de nosotros, de nuestros intereses y propósitos es sólo parcial. La razón por la que es sólo parcial es porque las funciones que se le asignan al indicador como instrumento están limitadas por lo que el instrumento *puede*

indicar; limitadas por aquello de lo que dependen sus varios estados y condiciones (Dreske 1988, 60).

De acuerdo con Dreske, la diferencia entre esta clase de sistemas representacionales, sistemas del tipo II, y los sistemas del tipo I es que mientras que en estos últimos tanto la función de sus elementos como el modo en que ellos realizan esa función derivan absolutamente de la capacidad de representación de otro sistema representacional, en los sistemas del tipo II la capacidad de representación de los elementos del sistema es intrínseca; no depende de mí o de ti o de nuestros intereses, ni de ningún otro ser con capacidad de representación, sino sólo de la constitución física del sistema respectivo. Lo que sí es derivativo en los sistemas del tipo II, lo que sí depende de nuestros intereses y de nuestros propósitos, es la determinación de la función que tienen sus elementos. Pero, de nuevo, la determinación de su función depende de las cosas que de hecho ya pueden realizar sus elementos.

Este es el momento para recordar que es fundamental encontrar una restricción que permita determinar cuál es el objeto representado por una representación, si es que se quiere o se puede seguir manteniendo la tesis de que la representación es básicamente un fenómeno causal. El problema que subyace a todos estos argumentos sobre las condiciones de buen funcionamiento de los mecanismos funcionales y las relaciones causales entre *representandum* y representación, es el de encontrar una respuesta a la "cuestión de la singularidad" (Stampe 1977, 44). En otras palabras, es responder satisfactoriamente al problema de la individuación de las representaciones en términos causales. El problema de la singularidad es el siguiente: ¿qué hace, pues, que sea un único objeto el objeto de una representación, y no otro cualquiera que esté relacionado causalmente con la producción de sus propiedades? La contestación de Stampe es que lo que hace de un único objeto el objeto de una representación es que, bajo

ciertas condiciones de buen funcionamiento de un mecanismo funcional, la especificación de sus funciones determina que sólo ese objeto, pero ningún otro más, cause las propiedades relevantes de la representación (1977, 44). La función de los mecanismos responsables de la etiología de las representaciones es fundamental para identificar la relación causal que subyace en la determinación de su contenido. Si somos capaces de identificar esta relación causal, habremos contestado a la cuestión de la singularidad.

Sin embargo, la determinación de las funciones del mecanismo depende, dentro de ciertos límites, de cómo queramos utilizarlo, de acuerdo con el ejemplo anterior del termómetro. Por consiguiente, no hemos respondido satisfactoriamente a la cuestión de la singularidad, porque, variando nuestros intereses, varían correspondientemente las funciones del mecanismo, y, así también, cuál sería la lectura correcta de su estado funcional.

De cualquier forma, apelar a la teleología de estos mecanismos funcionales es un peligro para las teorías causales de las representaciones. Parece que al recurrir a nociones teleológicas se está introduciendo subrepticamente la misma clase de nociones semánticas e intencionales que la teoría pretendía eliminar. Al menos para los sistemas representacionales del tipo II, el que un instrumento sirva para determinada función depende de que haya sido construido *con esa intención*. Pero en este caso las esperanzas que teníamos puestas en una teoría naturalista de las representaciones que no haga uso de nociones intencionales se van al traste.

& 6. Teorías causales *versus* teorías ortodoxas.

¿Son estas condiciones de buen funcionamiento de un mecanismo funcional más la especificación de sus funciones, condiciones necesarias y

suficientes para que una fotografía represente su objeto? Ciertamente son condiciones suficientes para que una fotografía sea una fotografía. Pero todos aquellos filósofos que Stampe denomina, no sin cierto tono crítico, "teóricos ortodoxos" argumentarían que una cosa es que algo sea la fotografía de un objeto y otra, bien distinta, que algo sea una representación (fotográfica) de un objeto. Por desgracia Stampe no da ningún nombre concreto cuando se refiere a los "teóricos ortodoxos". Sin embargo, el argumento que podría ofrecer uno cualquiera de ellos sería, muy probablemente, semejante a éste: Es verdad que la relación causal entre representación y *representandum*, más las condiciones óptimas de funcionamiento del mecanismo que media en la relación causal, son condiciones suficientes para que la fotografía sea una fotografía del objeto que causó sus propiedades bajo aquellas condiciones. No negamos sistemáticamente que en algunos casos haya una relación causal entre la representación y lo representado, y que, por esta razón, la representación tiene las propiedades que de hecho exhibe. Con todo, ser una representación (fotográfica) de un objeto no necesita de esta condición.

Los teóricos ortodoxos defienden la tesis de la distinción entre fotografías y representaciones (fotográficas) mediante un argumento que hace uso de la transitividad de las relaciones isomórficas entre propiedades. Si *R* representa a *O* porque hay una relación isomórfica estructural entre las propiedades *F* (*f*₁...*f*_n) del objeto *R* y las propiedades *O* (*O*₁...*O*_n) del objeto *O* y el objeto *P* tiene un conjunto de propiedades *P* (*p*₁...*p*_n) que están en una relación isomórfica estructural con las propiedades *O* (*O*₁...*O*_n) del objeto *O*—aunque no en una relación causal con él—, entonces la misma relación isomórfica se dará entre las propiedades *F* (*f*₁...*f*_n) del objeto *R* y las propiedades *P* (*p*₁...*p*_n) de *P*. Según este razonamiento, aunque sólo el objeto *O* haya causado las propiedades de *R*, *R* también representará a *P* en virtud del isomorfismo relevante de propiedades. Basta

con que lo decida yo así. Basta con que yo "aprehenda" esta relación isomórfica para que pueda *usar*, si así lo decido, el objeto *R* como una representación del objeto *P*.

La fotografía representaría, de acuerdo con los teóricos ortodoxos, a *G1* y a *G2*, porque podemos usar la fotografía para representar a ambos gemelos, aun cuando sólo *G2* haya sido quien ha causado las propiedades de la fotografía que son relevantes para que haya representación. Dado que *G1* y *G2* tienen muchas propiedades en común y cierto conjunto de propiedades de la fotografía tiene una relación isomórfica estructural con determinado conjunto de propiedades tanto de *G1* como de *G2*, podemos usar la fotografía para representar a *G1* y a *G2* y a todo otro objeto cuyas propiedades--o una parte relevante de ellas--tenga una relación isomórfica con las propiedades de la fotografía⁴. Dado que la relación isomórfica entre propiedades es transitiva, podemos usar la fotografía para representar a ambos gemelos. Pero si admitimos esta conclusión, entonces las relaciones causales entre *G2*, la cámara--en un estado de buen funcionamiento--y la fotografía, por una parte, y la relación de representación por la otra, serían relaciones diferentes, y no sería preciso explicar una en términos de la otra. Las relaciones causales no son *siempre* transitivas y las relaciones de representación por semejanza de propiedades sí lo son.

En el caso de las representaciones fotográficas, así como en otra clase de representaciones como, por ejemplo, las representaciones lingüísticas, los teóricos ortodoxos, de acuerdo con Stampe, defienden que "una cosa frecuentemente causa que su representación tenga las propiedades significativas que tiene (esto es, aquellas gracias a las cuales tiene el

⁴ Esta es la tesis que un teórico "ortodoxo" de las representaciones como Wittgenstein enuncia en el punto 2.121 del Tractatus, en donde dice: "Das Bild kann jede Wirklichkeit abbilden, deren Form es hat".

significado que tiene). Su posición simplemente es que ninguna relación causal semejante a ésta puede ser *esencial* para que la expresión se refiera al objeto al que se refiere" (Stampe 1977, 45). Mirando a la fotografía no podemos inferir, a partir de ella y de sus propiedades, que la fotografía es una fotografía de *G1* y no de *G2*. Mirando a la fotografía no podemos saber si fue *G1* o *G2* el que estaba enfrente de la cámara. Pero podemos afirmar que la fotografía representa (es decir, puede ser usada para representar) a los dos gemelos. La relación "ser una fotografía de" necesita de relaciones causales. La relación "ser una representación (fotográfica) de" sólo necesita de relaciones causales "indirectamente", por así decir. En este último caso sólo se necesita que un objeto haya causado alguna vez las propiedades de la fotografía, pero no precisamente el objeto respecto del cual usamos la fotografía como representación suya. La fotografía es una representación fotográfica tanto de *G1* como de *G2*, porque ciertos sistemas de representación--por ejemplo, los seres humanos--pueden usar la fotografía para representar a *G1* y a *G2*.

En resumen, los "teóricos ortodoxos" afirman que la representación no es esencialmente un fenómeno causal, aunque algunos procesos causales subyacentes sean necesarios para producir la representación. La representación es, sobre todo, un fenómeno relacionado con el "uso" o "aprehensión" que hagamos de ciertos objetos. Dicho clásicamente, en la producción de la representación es muy plausible que entren en juego, fundamentalmente, cadenas causales. Pero la especificación del contenido de una representación no es la determinación de la cadena causal que lleva del *representandum* a la representación. Desde este punto de vista, es la semejanza o isomorfismo de propiedades entre *G1*, *G2* y la fotografía, *conjuntamente* con el "uso" o "aprehensión" que hagamos de esa semejanza lo que es esencial para que haya representación (fotográfica).

Si no me equivoco, ésta es precisamente la tesis central de otro teórico "ortodoxo" de las representaciones: Husserl. A propósito de la crítica a la teoría que afirma que las "imágenes" de los objetos "externos" son sus representantes en la conciencia, Husserl hace las siguientes declaraciones, por lo demás completamente decisivas para el problema presente.

La semejanza entre dos objetos, por grande que sea, no hace del uno la imagen del otro. La imagen sólo se torna imagen por la facultad que un *yo* representante tiene de servirse de lo semejante como representante imaginativo de lo semejante, de tener presente intuitivamente sólo una cosa y *mentar* otra en lugar de ella. Pero esto sólo puede consistir en que la imagen se constituya como tal en una conciencia intencional peculiar, y en que el carácter *interno* de este acto, la peculiaridad *específica* de este "modo de apercepción", no sólo constituye lo que llamamos representar imaginativo, sino que constituye también lo que llamamos representación imaginativa de este o aquel objeto *determinado*, según la peculiar determinación, asimismo interna, del acto.... El cuadro sólo es imagen para una conciencia que constituya imágenes, la cual presta la *validez* o la *significación* de imagen a un objeto primario.... (Husserl 1984, 436-537).

El problema que entraña una tesis "ortodoxa" de este estilo es el de saber bajo qué condiciones el "préstamo de la validez" a un objeto como imagen o representación de otro está realmente justificado⁵. En el caso extremo, un teórico ortodoxo podría llegar a defender que habría representación (fotográfica), aunque las propiedades de la fotografía se hubieran producido por casualidad o accidente. Sólo la relación de semejanza y el uso o "aprehensión" de esa semejanza son condiciones estrictamente necesarias y suficientes, en este caso concreto, para que haya

⁵ Este problema ya se ha tratado con cierta minucia en el parágrafo 3 (b) del capítulo segundo, en donde se discutió la naturaleza de la relación signica entre la existencia de los fósiles y la existencia de animales antediluvianos.

representación, y no la relación causal entre *representandum* y representación.

Pero Stampe niega todas estas afirmaciones. Para él la intención no hace la representación (Stampe 1977, 46). Imagínese, dice, que G2 toma una fotografía a G1 con la intención de usarla como si fuera una fotografía de sí mismo. Esta intención no es suficiente, aunque es desde luego una condición necesaria, para que "de hecho" la fotografía sea una representación (fotográfica) de G2 (Stampe 1977, 46). No hay en este caso ningún control legal contrafactual entre las propiedades de la representación y G2. No podemos decir "la fotografía no habría tenido las propiedades que tiene si G2 no hubiese tenido las propiedades que tiene". (Para un ejemplo semejante véase Stampe 1977, 46).

Para explorar la plausibilidad de la tesis de Stampe y para que su contraargumentación no sea pura *ignoratio elenchi*, debemos preguntarnos qué quiere decir exactamente la afirmación "ortodoxa" de que podemos usar un objeto para representar a otro, aunque el primero no haya causado el segundo. Pudiera ser que, después de todo, haya casos en los que estaríamos dispuestos a decir que hay representación, aunque la tesis central de los teóricos ortodoxos no tenga ningún papel para explicar el fenómeno de la representación. En estas condiciones, deberíamos empezar a dar más credibilidad a las teorías causales. Lo que hay que demostrar es que hay circunstancias en las que se manifiesta una asimetría en la interpretación de las funciones que tienen los enunciados que expresan nuestras apprehensiones o usos de una imagen (fotográfica) como representación de un objeto, y que esta asimetría sólo se puede explicar si entran en juego los procesos causales que han originado la imagen (fotográfica).

Considérese el siguiente ejemplo ficticio. Un día estoy haciendo turismo y tomo una fotografía a una iglesia escondida en una calle poco conocida de cierta ciudad. Cuando revelo el carrete me encuentro con la

sorpesa desagradable de que la imagen que aparece en la fotografía se asemeja muchísimo--de parecerse a algo--a un cuadro abstracto, pero en absoluto a la iglesia que quise fotografiar. Mi cámara, me dicen en la tienda, funciona mal, y ésta ha sido la causa de que se haya echado a perder la fotografía. Esa misma tarde, cuando estoy visitando el Museo de Arte Contemporáneo de aquella ciudad, me encuentro, para sorpresa mía, con un cuadro que es completamente semejante a mi fotografía defectuosa de la iglesia. Algunos días más tarde, de regreso a casa, enseño la fotografía defectuosa de la iglesia a mis amigos y les digo: (i) "mirad, esta es una fotografía de un cuadro que vi el otro día en el museo".

Según este ejemplo y de acuerdo con el criterio que está buscando Stampe, es claro que el objeto que estoy enseñando a mis amigos no es una fotografía del cuadro que vi en el museo, aunque lo esté usando con esa intención. Es decir, la relación de semejanza entre la fotografía defectuosa de la iglesia y el cuadro abstracto, más la "aprehensión" de esa semejanza y la intención de usarla como fundamento de la relación de representación, no hacen que la fotografía sea una representación del cuadro, a no ser que haya una relación causal legal entre las propiedades del cuadro y las de la fotografía. En este caso es evidente que la fotografía habría tenido las propiedades que tiene, aunque el cuadro hubiera tenido otras propiedades distintas, y también aunque la iglesia hubiera tenido--como de hecho tiene--otras propiedades.

Si les pidiera a mis amigos que dieran una interpretación sofisticada del enunciado (i), es muy posible que ofrecieran una contestación que se expresaría en términos paralelos a éstos: las propiedades del cuadro habrían sido una condición suficiente para producir las propiedades de la fotografía si mi cámara hubiera funcionado óptimamente y el resto de las condiciones hubieran sido las mismas. En el ejemplo propuesto es evidente que la cláusula "si la cámara hubiera funcionado óptimamente" no se cumple, y,

en este sentido, la interpretación de mis amigos del enunciado (i) es falsa. Sin embargo, cuando yo enuncio (i), mi intención "suple" las condiciones de mal funcionamiento de la cámara. En esta circunstancia precisa, mi intención es una de las *causas* de que la fotografía defectuosa de la iglesia pueda ser usada por mí y posteriormente por mis amigos como una representación del cuadro. Bajo esta interpretación, no hay representación si no hay intención, y si es que esta intención no funciona como causa.

Si se analiza la función de las intenciones de los sujetos en esta clase de ejemplos, resalta el hecho de que estas intenciones son el resultado y expresión final de, al menos, la constatación y comparación del contenido de una serie de percepciones—las de la fotografía defectuosa y las del cuadro. Mi intención es la expresión del hecho de que yo sé cómo es la fotografía defectuosa, cómo es el cuadro y que hay una relación isomórfica entre algunas de sus propiedades. Así, pues, todos estos procesos de percepción y de comparación son causa de que yo pueda usar la fotografía como representación del cuadro. Mi intención, por su parte, es la constatación y expresión del contenido de estas percepciones y de su comparación, aunque mis amigos no lo sepan. Mis amigos, completamente ajenos a estas circunstancias, simplemente piensan que mi intención es la de expresar que yo hice una fotografía (buena) del cuadro.

Por consiguiente, en el enunciado mediante el cual expreso mi intención—esta fotografía defectuosa de la iglesia es una representación del cuadro abstracto—hay más información que la que hay contenida en la reconstrucción del significado del enunciado (i) por parte de mis interlocutores, lo sepan ellos o no. Yo dije a mis amigos: "ésta es una fotografía de un cuadro". La interpretación correcta del enunciado (i) necesita contar con el hecho de que este enunciado es la expresión del resultado de todas aquellas operaciones de comparación hechas por mí (mencionadas anteriormente); es decir, de mi conocimiento previo de cómo

es el cuadro y la fotografía defectuosa de la iglesia. Por el contrario, la información con la que cuentan mis amigos implícitamente es que la fotografía es simplemente una fotografía del cuadro, porque la cámara hizo lo que se suponía que tenía que hacer. En este sentido, se puede argumentar en favor de la teoría causal de las representaciones que es suficiente para la especificación del contenido representacional de la fotografía con determinar la cadena causal que lleva del *representandum* a la representación. En la mostración de la fotografía a mis amigos, éstos no hubieran pensado que esa fotografía es una representación del cuadro, si es que no hubiera existido la fotografía defectuosa de la iglesia y yo la hubiera comparado con el cuadro.

Considérese ahora el caso en el que mi cámara hubiera funcionado perfectamente bien y el resto de las condiciones fueran las mismas. Mi enunciado "mirad, esta es una fotografía de un cuadro que vi en el museo"—supongo que de hecho tomé una fotografía al cuadro—, es también la expresión de una intención que tengo. En esta ocasión hay una relación de semejanza entre la fotografía y el cuadro, una "aprehensión" de esa semejanza y la intención de usarla como fundamento de la aprehensión. Es decir, se dan todas y cada una de las condiciones de las que hablan los "teóricos ortodoxos". Pero en este ejemplo mi intención no es una de las causas de que la fotografía se pueda usar como una representación del cuadro. La expresión de esta intención aquí sólo afirma que la cadena causal entre el cuadro y la fotografía es correcta, óptima o relevante. Mi intención es la constatación de una cadena de hechos relacionados causalmente, y no un eslabón más en esa cadena. Es ajena, "exterior", a esa cadena causal. No es miembro de dicha cadena. La fotografía representa el cuadro independientemente de mi uso, de mis intenciones o de las de mis amigos. En esta variación del ejemplo, hay en mi enunciado la misma información o contenido que hay en la fotografía.

¿A dónde quiero ir a parar con estas reflexiones? A la afirmación de que la tesis de Stampe según la cual no hay representación si no hay causación, y que la representación es un fenómeno causal, empieza a ser, cuando menos, plausible. En aquellos casos en los que la intención, el "uso" o la "aprehensión" parecen jugar un papel predominante en la explicación de las representaciones y de sus contenidos, nos encontramos con que las intenciones mismas son causas. Si tanto en el caso de la fotografía defectuosa de la iglesia como en el de la fotografía "buena" del cuadro los teóricos ortodoxos afirman que la relación de semejanza entre el cuadro y la fotografía y la aprehensión de esa relación como fundamento de la representación, son condiciones necesarias y suficientes para que haya representación, entonces queda sin explicación alguna la intuición de que si la cámara hubiera funcionado bien, la fotografía habría representado el cuadro independientemente de mis intenciones y la de mis amigos. En otras palabras, se queda sin explicación la diferencia en la interpretación del enunciado (i). Si alguien tuviese alguna razón de principio en contra de la posibilidad de que la teoría causal de las representaciones pueda proporcionar condiciones suficientes para especificar el contenido de una representación, debería preocuparle este ejemplo.

La fotografía defectuosa no representaría el cuadro si las propiedades de mis intenciones no estuvieran controladas contrafácticamente por las propiedades del cuadro y la fotografía. En otras palabras, si mis intenciones no fueran o no funcionaran como un mecanismo con rendimiento o funcionamiento óptimo, la fotografía no podría ser una representación del cuadro. Los mecanismos de la intención suplen a los mecanismos de una cámara que funcione bien.

¿Qué conclusiones hay que extraer de todos estos ejemplos? ¿Qué hay que decir de la teoría de Stampe? ¿Cuáles son sus problemas

fundamentales? Según mi opinión, y a la luz de las consideraciones precedentes, los problemas más importantes son tres.

(1). El problema de la teleología. La especificación de las funciones del mecanismo depende, dentro de ciertos límites, de nuestras intenciones y propósitos. Por consiguiente, no hemos conseguido todavía una teoría que dé cuenta de los fenómenos representacionales en términos naturalistas, esto es, en términos que no sean ni semánticos ni intencionales.

(2) El problema de la individuación del contenido de las representaciones, que está íntimamente relacionado, como se ha visto a lo largo de la presente discusión, con el punto anterior.

(3). El problema de la distinción entre la verdad y la etiología de las representaciones. Bajo condiciones de funcionamiento óptimas de los sistemas representacionales del tipo II no hay posibilidad de distinguir entre las condiciones--tanto internas del mecanismo como externas del entorno--responsables de la etiología de la representación y sus condiciones de verdad. Si *O* causa *R* es porque (a) *O* debe ser contado en el dominio de los objetos que pueden causar, y (b) porque si *O* causa *R*, entonces es que *O* es el caso. Esta es, según Fodor, la debilidad principal de las teorías causales de las representaciones.

Supóngase que *S* es la condición de verdad de *R* en razón del hecho de que es la causa de *R*. Ahora bien, la causación es diferente de la semejanza de la siguiente manera. Un símbolo puede, supongo yo, ser semejante a algo meramente posible. Está muy bien que un cuadro sea el cuadro de un unicornio. Pero, seguramente, un símbolo no puede ser el efecto de algo meramente posible. Si *S* causa *R*, entonces *S* es el caso. Pero si *S* es el caso y *S* es la condición de verdad de *R*, parece como si *R* tuviera que ser verdad si es que el ser verdad es tener condiciones de verdad que son el caso. Así que parece que una teoría que cuenta la *causación* entre las relaciones en virtud de las cuales una representación tiene sus condiciones de verdad permitiría que se asignaran condiciones de verdad sólo cuando se satisfacen (1984b, 238-239).

Parece, por tanto, que tampoco es del todo satisfactoria una teoría ortodoxa de las representaciones por las razones expuestas anteriormente. ¿Qué camino se puede seguir a partir de ahora?

& 7. Sistemas representacionales del tipo III.

Quizá todas estas objeciones y todos estos problemas tengan su origen en el hecho de que en los mecanismos que hemos estado analizando somos nosotros quienes, hasta cierto punto, determinamos sus funciones y, correspondientemente, qué representan. Sin embargo, no parece que sea éste el caso con otra clase de sistemas de representación. Hechos tales como que las huellas digitales, el canto de los pájaros, los fósiles, los anillos en el tronco de un árbol, etcétera, representen, respectivamente, a una cierta persona, una ceremonia de aparejamiento, la existencia de animales antediluvianos o la edad del árbol, no parece que dependan de que nosotros les atribuyamos o no estas funciones. A primera vista parece como si todos estos estados de cosas indicaran lo que de hecho indican, sin que nosotros intervengamos para nada en su determinación. Hasta donde *nuestros conocimientos de biología alcanzan*, parece ser cierto que el número de anillos en el tronco de un árbol representa o indica la edad del árbol. Con esta afirmación no quiero dar a entender que el número de anillos en el árbol represente o indique la edad del árbol porque nosotros tenemos una teoría sobre esta clase de organismos vivientes que nos permite hacer esta aseveración. Estoy dispuesto a creer que el canto de los pájaros, los fósiles y los anillos en el tronco de un árbol indican lo que indican independientemente de lo que nosotros seamos capaces de decir con verdad sobre ellos, o de cómo los utilicemos para seguir progresando en la investigación científica. Más aún, independientemente de si somos capaces o no de reconocer su significado (Dreske 1988, 54). En

todos estos casos, que normalmente se conocen con el nombre de "signos naturales", ciertos estados de cosas "derivan su capacidad para indicar no (como en el caso de los signos indicativos) de nosotros, del uso que hacemos de ellos para indicar, sino del modo en que ellos están relacionados objetivamente con las condiciones que significan" (Dreske 1988, 54). La forma en que están relacionados el estado de cosas que indica y el estado de cosas que es indicado, es, muy frecuentemente, causal.

Dreske denomina esta nueva clase de sistemas representacionales "sistemas representacionales del Tipo III" (1988, 62). Las dos características fundamentales de estos sistemas de representación son que tanto sus funciones como el modo en el que sus elementos expresivos realizan estas funciones no dependen de ningún otro sistema de representación ni de cómo se exploten cognoscitivamente estas funciones. En estos casos se dice que la capacidad de representación es *intrínseca* al sistema.

Según estas primeras consideraciones, parece también claro que el número de anillos en el tronco de un árbol no representa necesariamente de una manera precisa la edad del árbol. De acuerdo con nuestras teorías biológicas, en años de sequía severa puede suceder que el árbol no produzca un nuevo anillo y, como consecuencia, que no haya una correspondencia unívoca entre el número de anillos y la edad real del árbol. Es muy probable que las células responsables del crecimiento del árbol sólo cumplan apropiadamente sus funciones en determinadas condiciones locales "normales", y que en otras condiciones, como podrían ser años de sequía, los anillos del árbol, de los cuales son responsables causalmente aquellas condiciones, no sirven como indicadores de la edad del árbol.

Como dije anteriormente, la función de indicación de los anillos en el tronco del árbol parece ser independiente de cómo usemos estos signos naturales. Sin embargo, aun cuando nuestras teorías biológicas en boga sean

parcial o totalmente falsas o incompletas, nos permiten emitir hipótesis sobre cómo interpretar correctamente los anillos en el tronco del árbol. En concreto, nos permiten decir qué hubiera causado la producción de un nuevo anillo. Si se hubieran dado ciertas circunstancias "normales" en el entorno local del organismo, entonces habría aparecido un anillo más después de haber transcurrido un año. Estas circunstancias son condiciones climáticas, condiciones biológicas del árbol, condiciones de humedad y composición química del suelo, etc. Si estas condiciones se dieran, entonces se produciría la aparición de un nuevo anillo cuando la Tierra completara una revolución alrededor del Sol. En este caso, el número de anillos representaría la edad del árbol bajo el supuesto de que se dieran aquellas condiciones "normales". En condiciones normales no habría lugar para algo así como una indicación falsa.

Como se puede comprender fácilmente, éste es el mismo problema que teníamos en el ejemplo de las fotografías defectuosas y "buenas". Bajo condiciones "normales", no hay lugar para las representaciones falsas (misrepresentations). Explicar cómo es posible que haya representaciones falsas es un problema que, como han señalado repetidamente Schiffer y Fodor –a mi modo de ver, correctamente–, parece ser inherente a toda teoría causal sobre el contenido de las representaciones. Si las condiciones para que haya representación son las que Dennis Stampe ha establecido, entonces esas mismas condiciones son también condiciones para la verdad de la representación. Si se satisfacen las condiciones para que haya representación, entonces no puede haber representaciones falsas (Fodor 1984b, 234; Schiffer 1989, 47).

Para interpretar los anillos en el tronco del árbol, lo que hacemos es proponer hipótesis acerca de la causa de la representación.

Es una hipótesis *condicional*: que si ciertas condiciones tienen lugar <condiciones climáticas, condiciones del suelo, condiciones de las

células del árbol...>, entonces algo que tuviera tales y cuales propiedades <por ejemplo, cierta temperatura del microclima> causaría que la representación tuviera tales y cuales propiedades. Aun cuando esta lectura identifica a ciertos estados de cosas que causarían la representación en el supuesto de que se dieran ciertas condiciones, adviértase que no identifica éste como el único estado de cosas que causaría que el tronco tuviera ese número de anillos, incluso bajo aquellas condiciones normales. *Incluso bajo aquellas condiciones normales*, puede haber otras cosas que produzcan los anillos--puede ser que un ejército de alguna especie de hormigas devoradoras o un genio maligno omnipotente de los árboles. (Stampe 1977, 49). (El subrayado es mío).

A primera vista, esta tesis tiene dos interpretaciones posibles. Desde luego, es ciertamente plausible pensar que el conjunto de condiciones normales--¿condiciones en el entorno natural del organismo?--que nuestras teorías biológicas son capaces de recoger en el estado actual del conocimiento no sean *todas* las condiciones "normales" que de hecho han causado la producción anual de un nuevo anillo. Los estados de cosas de los que afirmamos que son la causa de la representación son sólo, quizá, una parte de las condiciones globales que se deberían dar en la realidad--y que deberían recogerse en la cláusula condicional de la hipótesis--para que se produzca la representación. Quizá haya más condiciones normales que han cooperado en la producción de los anillos en el tronco del árbol. Pero si las condiciones normales son *todas* las condiciones normales que se requieren para la producción de un nuevo anillo y es verdad que estas condiciones son suficientes para su producción, es difícil imaginar bajo qué condiciones otros estados de cosas, como un genio maligno de los árboles todopoderoso, pudieran causar el mismo efecto "incluso bajo aquellas condiciones normales".

La segunda interpretación viene exigida, precisamente, por la necesidad de superar esta dificultad. Incluso bajo la condición de que se den aquellas circunstancias normales, el número de anillos en el tronco del árbol se

puede producir accidentalmente, por casualidad, por un ejército de hormigas devoradoras; de tal forma que el número de anillos en el árbol representaría con exactitud (accurately) la edad del árbol bajo aquellas condiciones accidentales. Esta interpretación se acomoda perfectamente con otra de las tesis fundamentales de Stampe.

Las representaciones pueden ser exactas (accurate) completamente por casualidad o accidente, del mismo modo que los enunciados pueden ser verdaderos meramente por casualidad. (1977, 50).

Sin embargo, creo que se puede argumentar que los sistemas de representación del tipo III tienen ciertas propiedades que nos pueden llevar a pensar que esta tesis es, cuanto menos, cuestionable bajo una cierta interpretación.

Me parece que el problema que se plantea es el siguiente: no es fácil entender, aunque no parezca impensable, cómo es posible que haya una cadena causal que sea suficiente para la producción de un efecto determinado—las condiciones normales para la causación de un nuevo anillo, por ejemplo—, y que, al mismo tiempo, se dé otra cadena causal que sea ella misma también suficiente para la causación del mismo efecto—el genio maligno de los árboles todopoderoso que causa la producción del mismo número de anillos—, de forma tal que conservándose todos los eslabones de la primera cadena sea la segunda la que de hecho ha causado el efecto. Para que esta segunda posibilidad se dé, es preciso pensar que la primera cadena no era suficiente para la causación del efecto, en contra de la hipótesis de la que se partía. Si las dos cadenas causales son suficientes para la producción de un determinado efecto y las dos se dan al mismo tiempo, entonces habría algo así como una "sobredeterminación" del efecto (Schiffer 1989, 148).

Una consecuencia de esta reflexión es que si las condiciones normales son suficientes para la causación del efecto y, por consiguiente, son suficientes para que se dé una representación exacta o precisa, entonces no se pueden dar representaciones exactas por casualidad o accidente bajo el supuesto de que se den aquellas condiciones normales suficientes. Considérese el siguiente ejemplo ficticio, en el que se ignora por principio las dificultades derivadas del problema de la "sobredeterminación" del efecto. Imagínese que un biólogo trabaja en la investigación de las condiciones que favorecen el crecimiento de los árboles y que, por alguna razón, deja este proyecto y abandona su laboratorio para siempre. El laboratorio está en un lugar del planeta donde hay terremotos con cierta frecuencia, de tal forma que las vibraciones de las estanterías del laboratorio producen que se rompan ciertas probetas que contienen una sustancia química muy especial. Por casualidad, estas probetas vierten su contenido sobre un árbol próximo y producen la aparición en él de un nuevo anillo cada año. Según nuestra hipótesis, antes de que se rompiesen las probetas y de que vertieran su contenido sobre el árbol, se daban ya ciertas condiciones normales—el suelo tenía suficiente agua, la temperatura en el laboratorio era la adecuada, etc.—que eran suficientes para causar "naturalmente" un nuevo anillo. Sin embargo, lo que causó "de hecho" la aparición de un nuevo anillo fue el efecto de estas sustancias químicas que cayeron por casualidad, por accidente, sobre el árbol. ¿Significa esto que el nuevo anillo en el tronco del árbol es producido por casualidad o que los anillos representan por casualidad la edad del árbol? No lo creo así. En este ejemplo, las condiciones responsables del crecimiento del árbol habrían causado de todas formas la aparición de un nuevo anillo, aunque no hubieran sido estimuladas por aquellas sustancias químicas.

El resultado de este ejemplo es que las representaciones que llamamos enunciados pueden ser, quizá, verdaderas o precisas por casualidad⁶; pero una representación no puede ser precisa por casualidad en el caso en el que ciertas condiciones normales sean reponsables causalmente de la aparición de los elementos que tienen naturalmente la función de indicar un estado de cosas en un sistema representacional del tipo III.

Por supuesto, el número de anillos en el tronco del árbol puede representar de una forma imprecisa (falsa) la edad del árbol. Sin embargo, el hecho de que la representación sea imprecisa (inaccurate) no depende de lo que constituye una interpretación correcta de la representación (Stampe 1977, 49). Según la teoría de Stampe, una interpretación es correcta cuando deriva o depende de una generalización causal *verdadera* justificada contrafácticamente. Si determinadas condiciones del microclima se dieran, y los diferentes sistemas biológicos del árbol funcionaran propiamente..., entonces se formaría un nuevo anillo en el tronco del árbol con cada una de las rotaciones de la Tierra alrededor del Sol. A esta idea añade Stampe la siguiente:

Pero la lectura o interpretación correcta es distinta de y puede que no proporcione la comprensión causal correcta de las propiedades del tronco del árbol. (Stampe 1977, 49).

Estoy plenamente de acuerdo con Stampe, si esta ultima tesis significa que la interpretación adecuada de las propiedades del tronco del árbol--en este caso, el número de anillos--no está en una correlación necesaria con el conocimiento de los factores externos que de hecho han causado la producción de los anillos. La hipótesis de que las condiciones normales del entorno han causado la producción de un anillo más en el

⁶ Las ilustraciones típicas de este caso son las de una frase "escrita" por las olas en la arena de la playa, o por el viento en las dunas del desierto.

tronco puede ser errónea. En el ejemplo ficticio anterior fueron ciertas sustancias químicas las que produjeron causalmente la aparición del anillo. En este caso concreto, por consiguiente, una generalización contrafáctica como ésta: "si ciertas condiciones del microclima se dieran y los sistemas biológicos del árbol funcionaran propiamente..., entonces se formaría un nuevo anillo en el tronco del árbol con cada una de las rotaciones de la Tierra alrededor del Sol" no ofrece ninguna interpretación causal correcta de la aparición de los anillos en el tronco. La interpretación causal correcta de las propiedades del tronco es, más bien, esta otra: "si ciertas sustancias químicas con tales y cuales propiedades están en contacto con los árboles y los sistemas biológicos del árbol funcionan propiamente y..., entonces se formaría un nuevo anillo en el tronco del árbol".

Se pueden hacer diferentes generalizaciones causales justificadas contrafácticamente sobre la existencia de un efecto concreto; pero determinar cuál de entre ellas es la verdadera en cada caso no se puede establecer, por así decir, atendiendo exclusivamente a lo que dice cada una de las cláusulas condicionales. Hay que recordar una vez más que, según esta teoría, la especificación del contenido de una representación es un enunciado que expresa qué habría causado la representación en condiciones de fidelidad. En cada caso concreto sólo una especificación es la verdadera. O el ejército de hormigas devoradoras o las condiciones "normales" o el genio maligno de los árboles. Pero entonces se plantea un nuevo problema. Si diferentes generalizaciones contrafactuales pueden ser verdaderas y, por lo tanto, todas ellas expresan condiciones suficientes para la especificación del contenido de una representación, entonces parece que cada representación expresaría su contenido *más* las condiciones responsables de su etiología. Sin embargo, la idea intuitiva es que los anillos del árbol sólo representan su edad, y no que representan los anillos del árbol o las condiciones de su producción (Fodor 1984b, 243-244).

Pero éste no es el único problema. ¿Qué quiere decir Stampe con la afirmación de que una interpretación es adecuada si depende de una generalización contrafáctica *verdadera*? Me parece a mí que algo así como que una generalización contrafáctica, si verdadera, es una interpretación adecuada que nos proporciona una comprensión causal correcta de las propiedades de la representación –de los anillos del árbol.

Como he indicado hace un momento, es muy plausible que los elementos que funcionan como indicaciones o representaciones en los sistemas naturales de representación, indiquen o representen precisa o exactamente un cierto estado de cosas por casualidad. Es posible concebir la posibilidad de que el número de anillos en el árbol represente la edad del árbol, aunque se hayan dado varios años de sequía severa. Simplemente el genio maligno todopoderoso de los árboles ha causado el número exacto de anillos. En este sentido, Stampe tiene toda la razón cuando indica que la propiedad que tiene una representación de ser precisa no se puede identificar con la propiedad de que se "haya producido bajo cierto conjunto de condiciones, o con la propiedad de tener una cierta causa" (1977, 50).

Pero lo que no puede ocurrir, o al menos ésa es mi tesis, es que si una interpretación es correcta–si se dieran ciertas condiciones del microclima y el suelo tuviera tal grado de humedad, etcétera–, entonces los elementos que funcionan como representaciones en los sistemas naturales de representación indicaran de una forma precisa un cierto estado de cosas por casualidad. Así, por ejemplo, los anillos la edad del árbol. Una vez más Fodor y Schiffer tienen razón: no puede haber representaciones falsas en estas condiciones. A esta afirmación añado esta otra: ni verdaderas por casualidad en condiciones óptimas. Dreske expresa esta misma objeción cuando dice que "nada puede significar que P en el sentido natural si P no es el caso" (Dreske 1988, 55).

& 8. Las representaciones lingüísticas.

El fin último que persigue Stampe con su teoría causal de las representaciones es desarrollar una teoría naturalista de las representaciones lingüísticas. En concreto, piensa que se pueden explicar las representaciones lingüísticas sin recurrir a términos semánticos tales como "referencia" o "intención comunicativa de los hablantes". Para explicar el contenido de las representaciones lingüísticas deberían ser suficientes las nociones de "relación causal", "condiciones de buen funcionamiento de ciertos mecanismos" y otras semejantes, si es que es posible construir una teoría naturalista de esta clase de representaciones. En lo que sigue, me centraré en su teoría de los nombres propios, porque tiene un interés especial para entender posteriormente la crítica de Fodor a la teoría de Stampe, así como la que ha propuesto Fodor para superar los problemas que han ido apareciendo hasta este momento.

La tesis principal de Stampe es que hay ciertos mecanismos que gobiernan la introducción de nombres nuevos en el lenguaje. Si estos mecanismos funcionaran óptimamente, entonces no se introducirían en el lenguaje nombres vacuos. Si hay representaciones sin objeto, entonces es que estos mecanismos no han funcionado bien.

Una vez más, el problema que se plantea en el caso de las representaciones lingüísticas es el de explicar cómo es posible que haya representaciones falsas (misrepresentations) si se tienen en cuenta las restricciones que Stampe impone a la hora de especificar en términos naturalistas el contenido de cualquier representación en general. Una de las características fundamentales de la intencionalidad de las actitudes proposicionales es su capacidad para representar falsamente. Se puede creer *que P* cuando *P* no es el caso. Si la intencionalidad es un rasgo esencial

de las actitudes proposicionales, es claro que se puede decir o significar *que P* cuando *P* no es el caso (Dreske 1988, 64-65).

Los nombres vacuos son aquellos que no indican o designan objeto existente alguno. También para Stampe, los nombres propios, en oposición a otras representaciones lingüísticas, no tienen significado, y, consecuentemente, no son en sí mismos representaciones de ningún objeto. Pero están íntimamente asociados con representaciones. El ejemplo que ofrece Stampe de un nombre vacuo es "Vulcano". Una representación--en este caso una descripción definida-- que se asocia, o que se asoció en un momento histórico determinado, con este nombre es ésta: "El planeta con una órbita intramercurial responsable del cambio de dirección del perihelio de Mercurio". El nombre es vacuo, porque no hay semejante objeto. La vacuidad de la representación que se asocia con el nombre propio causa o puede llegar a causar que el nombre tenga también esta propiedad.

De todas las maneras, aunque el nombre sea vacuo, asumimos que algo es responsable del cambio de dirección del perihelio de Mercurio. También podemos asumir que si entendemos el significado de la descripción que se asocia con el nombre "Vulcano", entonces podemos aprender algo (equivocado) acerca de la causa del cambio de dirección del perihelio de Mercurio.

De acuerdo con la teoría de las descripciones de Russell, la descripción que se asocia con el nombre "Vulcano" se puede parafrasear diciendo que "una y solamente una entidad es responsable del cambio de dirección del perihelio de Mercurio, y esa entidad es un planeta con una órbita intramercurial". Si analizamos esta paráfrasis russelliana de la descripción originaria se manifiesta que el segundo miembro de la conjunción no designa ningún objeto. El predicado "planeta con una órbita intramercurial" sencillamente no es verdadero de ningún objeto. Por el contrario, el enunciado abierto o no ligado--una y sólo una entidad...--

designa la causa actual del cambio de dirección del perihelio de Mercurio, a saber, la curvatura del espacio/tiempo en la proximidad del Sol. De acuerdo con Stampe, el resultado de combinar estas dos partes es una representación falsa (misrepresentation), *pero representación al fin y al cabo*, de la causa actual del cambio de dirección del perihelio de Mercurio. La prueba para determinar que la descripción "el planeta con una órbita intramercurial..." es una representación falsa, *pero representación al fin y al cabo*, de la causa real de este fenómeno --el cambio de dirección del perihelio de Mercurio-- consiste en averiguar si se da el necesario control contrafactual entre la causa actual de este fenómeno y la representación falsa. Pero tal control existe.

Si la curvatura del espacio fuera mayor de lo que de hecho es, entonces ese enunciado <"Vulcano describe una órbita a dieciocho millones de millas del Sol"> acerca de la distancia de Vulcano al Sol, habría variado correspondientemente (Stampe 1977, 57).

La conclusión que extrae Stampe de la existencia de este control contrafactual es que la descripción que se asocia con el nombre "Vulcano" es una descripción inexacta (inaccurate) de la curvatura del espacio/tiempo en la proximidad del Sol, porque es una representación falsa de la causa real del cambio de dirección del perihelio de Mercurio.

De acuerdo con este filósofo, es necesario distinguir entre el nombre "Vulcano", la curvatura del espacio/tiempo en las proximidades del Sol y la representación falsa (*pero representación al fin y al cabo*), de la curvatura del espacio/tiempo. La representación falsa, *pero representación al fin y al cabo*, que se asocia con el nombre "Vulcano", hace, en este caso concreto, que el nombre sea vacuo.

Sin embargo, Stampe afirma que el resultado de asociar una representación falsa con un nombre no siempre es un nombre vacuo. En

el caso del nombre "átomo" es claro que muy diferentes representaciones se han asociado con él en el transcurso del desarrollo de nuestras teorías sobre el mundo físico. Con todo, Stampe dice que el nombre "átomo" siempre se refiere a lo mismo (al mismo objeto), aunque se haya asociado con aquel nombre una serie de descripciones más o menos precisas a lo largo del tiempo.

Pero entonces, ¿cuál es el criterio para distinguir entre "Vulcano" y "átomo"? Lo que establece la línea divisoria entre estos dos nombres es que en el caso de "Vulcano" no se dio el conjunto relevante de condiciones óptimas responsables de su introducción en el léxico científico, mientras que no ocurrió lo mismo en el segundo caso. ¿Cuáles son, pues, estas condiciones óptimas, según Stampe?

En el caso de los nombres vacuos, como "Vulcano", la idea es ésta: una cierta parte del teorizar científico crea la representación de un cierto fenómeno. Pero sucede que es una representación falsa (misrepresentation) de ese fenómeno, esto es, de la causa del cambio de dirección del perihelio. Adviértase que lo que digo es que ésta era de hecho una representación de algo real. Podemos concebir esta representación como el producto de ciertos artificios cognitivos (la maquinaria inductiva y abductiva) y decir que, si se hubieran dado las condiciones para la manipulación exitosa del artificio, la representación que en ese caso se hubiera producido habría sido una representación precisa de la causa del cambio de dirección del perihelio—esto es, habría sido así en el caso de que los artificios en cuestión sean competentes para realizar sus funciones. (Stampe 1977, 58).

Stampe considera necesario que las descripciones que se asocian con el nombre "Vulcano" representen precisamente, adecuadamente, la causal real del cambio de dirección del perihelio (1977, 58). Solamente en el caso de que estas representaciones fueran adecuadas o precisas, el nombre "Vulcano" hubiera estado libre del peligro de ser un nombre vacío. Esta condición, sin embargo, no es necesaria para el nombre "átomo", como tampoco para

las descripciones que se asocian con él. Manifiestamente, surge de nuevo el problema de saber por qué estas representaciones tienen que ser adecuadas o precisas en el caso de "Vulcano" pero no en el caso de "átomo". Para expresar la misma cuestión con otras palabras: ¿qué significa decir que los mecanismos inductivos y abductivos funcionaron lo suficientemente bien cuando se produjo el nombre "átomo"--de forma tal que "átomo" representa sólo imprecisamente, pero no falsamente--, y lo suficientemente mal cuando se introdujo el nombre "Vulcano" --de tal forma que "Vulcano" representa falsamente?

Lo primero de todo, merece la pena saber cuáles son las funciones de estos mecanismos inductivos y abductivos. Su función principal es alcanzar la verdad. Dicho con palabras de Stampe, su función es "restringir las innovaciones terminológicas y prevenir la proliferación innecesaria y disfuncional de términos. Quizás funcionen de una forma tal que, si fuera óptima, produciría y asignaría un nombre a cada clase notable de objeto (natural), a cada persona, etc., y *sólo un* nombre" (1977, 58).

El verdadero problema con esta tesis es que estas restricciones son demasiado severas para introducir un nuevo nombre en el vocabulario científico y, en general, para introducir un nombre en el lenguaje. Si aceptamos estas restricciones, el mecanismo responsable de la introducción de los nombres sería tal que funcionaría óptimamente sólo en el caso de que tuviéramos un conocimiento adecuado o preciso del mundo físico. O, incluso, bajo otra condición todavía más severa, a saber: bajo el supuesto de que tuviéramos un solo nombre para cada objeto realmente existente. En otras palabras, si tuviéramos un conocimiento perfecto y preciso de cada clase natural relevante de objeto--la clase de conocimiento que se encuentra en el límite de nuestras aspiraciones cognoscitivas y que sólo puede tener un ser que posea un conocimiento perfecto--, entonces no se introducirían nombres vacíos. Pero si estas restricciones se pusieran en

práctica, la ciencia sería casi imposible, porque, en muchos casos, sólo tenemos un conocimiento provisional del mundo físico, y en muy pocas ocasiones estaríamos completamente seguros de si las descripciones que asociamos con un nombre representan adecuadamente el objeto que describen.

Ciertamente, tiene que haber un método que restrinja la introducción en el lenguaje de demasiados nombres o de demasiado pocos. Pero no creo que este mecanismo se pueda o deba identificar con estas condiciones de buen funcionamiento de los mecanismos inductivos y abductivos *à la* Pierce de los que habla Stampe. Además, ¿qué sería un mecanismo inductivo y abductivo perfecto? Sería poseer algo así como todos los enunciados observacionales verdaderos de cada región del ser y las hipótesis verdaderas que expliquen sus relaciones nomológicas.

¿Cuál es, entonces, el mecanismo que gobierna la introducción de nuevos nombres, dado que alguno tiene que haber? Pensemos de nuevo en la analogía entre los nombres propios y las etiquetas de los archivos, para entender cómo se lleva a cabo la introducción de nuevos nombres en el lenguaje. Si tenemos cierta cantidad de información--quizá en la forma de descripciones relacionadas con un cierto fenómeno--y pensamos que esta información merece la pena de ser almacenada por diversos motivos, entonces creamos un archivo en nuestra mente que contenga esta información y le damos un nombre. Los nombres, en este sentido, no tienen significado, sólo son marcas para distinguir unos archivos de otros. El mecanismo que crea archivos y los nombra funciona con las siguientes limitaciones generales. Por una parte, no es conveniente poner todos los documentos o datos que tenemos en un único archivo, porque entonces tendríamos demasiado pocos nombres y no conseguiríamos una clasificación útil y productiva de la información ni, en última instancia, del mundo. Por otra parte, no es conveniente crear un archivo para cada uno de los datos

o documentos, por la misma razón. Pero, en cualquier caso, no necesitamos esperar a tener una clasificación perfecta de todos los datos para abrir un nuevo archivo. Estamos perfectamente legitimados a abrir archivos aun cuando no sepamos si la información que ponemos en ellos va a ser por fin almacenada de la forma en que ahora lo está. Quizá, cuando el tiempo pase, nos demos cuenta de que todos o alguno de los datos de este archivo deberían ser extraídos de él y colocados en otro archivo. También es muy posible que de vez en cuando necesitemos añadir nuevos datos en este archivo.

Con el desarrollo de las teorías científicas realizamos continuos cambios de información de archivo a archivo. En el caso del nombre/archivo "átomo", siempre se mantiene una reserva de datos en él, aunque la información relevante (o parte de ella) haya cambiado sustancialmente desde que, digamos, Demócrito de Abdera y Leucipo lo crearon. En el caso del archivo "Vulcano", sencillamente no hay razón para conservarlo. Es mejor sacar los datos que teníamos en él y ponerlos en otro archivo o, quizá, deshacernos de ellos para siempre. Un nombre vacío es el nombre de un archivo que está vacío o que debería estarlo. La maquinaria inductiva y abductiva nos puede decir cuándo necesitamos añadir o quitar datos a los archivos. Pero, por lo que sabemos, aun cuando esta maquinaria inductiva y abductiva esté funcionando "óptimamente", es posible que se introduzca un nombre vacío. Con el trasfondo cognitivo de la teoría newtoniana sobre el mundo físico, Leverrier tenía cierta información útil que deseaba conservar. Por ejemplo, sabía que había un cambio de dirección en el perihelio de Mercurio. Por otra parte, de acuerdo con su teoría y los mecanismos abductivos que ella ponía a su disposición, ¿qué otra cosa podría causar el cambio de dirección, si no era un planeta hasta entonces no observado? También pensaba que era deseable almacenar esta información en un archivo que se llamara "Vulcano". Eso es lo que hizo.

El resultado de estas reflexiones es que no es necesario que los datos de un archivo contengan información adecuada o exacta para que se cree legítimamente. Sólo una secretaria completamente eficiente puede almacenar perfectamente la información en archivos y etiquetarlos con nombres de forma que no se requieran más cambios en el futuro.

Tomando como punto de partida estas reflexiones, creo que Fodor tiene razón cuando critica la teoría de las representaciones lingüísticas de Stampe y afirma que los mecanismos que fijan las creencias, aquellos que hacen que tengamos una creencia con un contenido determinado, pueden estar funcionando perfectamente bien aun cuando la información que proporcionen sea falsa o inadecuada (Fodor 1987, 105). Puede haber mecanismos que funcionen perfectamente bien y que fijen por sistema creencias falsas porque han sido diseñados para cumplir esa función. El modelo de la creación de archivos para explicar la introducción de nuevos nombres proporciona una analogía interesante para hacer plausible esta crítica. Pero antes, una advertencia. Es cierto que Stampe sólo habla de mecanismos responsables de la introducción de nuevos nombres, mientras que Fodor habla acerca de los mecanismos responsables de la fijación de las creencias. Sin embargo, dado que las restricciones que impone Stampe para que se introduzca un nuevo nombre son tan severas, la relación entre las condiciones de buen funcionamiento de los mecanismos inductivos y abductivos que generan nuevos nombres y los mecanismos para generar creencias verdaderas, es evidente y manifiesta. La clasificación eterna y perfecta de los objetos naturales de la manera que se ha esbozado más arriba exige tener creencias verdaderas sobre estos objetos. Sólo las creencias verdaderas proporcionan representaciones adecuadas que eviten la introducción de nombres vacuos. Todo esto, naturalmente, bajo el supuesto de un mundo real y sistemático, exterior a nuestras percepciones.

¿Puede haber mecanismos que funcionen óptimamente y que fijen o rindan creencias falsas? Hasta ahora he estado argumentando que la asociación entre las condiciones de funcionamiento óptimas de los mecanismos inductivos y abductivos con la clase de conocimiento que tendría un ser cognitivamente perfecto, es muy estrecha. Aunque es difícil imaginar en qué consistiría un mecanismo inductivo perfecto--¿un mecanismo que rindiera todos los enunciados observacionales verdaderos posibles?--, la analogía con los archivos que he hecho antes sugiere que las condiciones óptimas para crear nuevos archivos y el hecho de tener información adecuada en estos archivos pueden ir en direcciones contrarias. No podemos identificar los nombres vacíos apelando a las circunstancias óptimas de funcionamiento de los mecanismos cognitivos sin haber identificado antes, de una forma completamente independiente, estas circunstancias óptimas como aquellas que rinden descripciones o creencias verdaderas. Al menos no podemos hacerlo, si es que queremos definir "las *condiciones* de verdad como aquellas que se satisfacen cuando un símbolo (en este caso, un nombre propio) es causado en circunstancias óptimas" (Fodor 1987, 105). Si lo hiciéramos, estaríamos usando la misma clase de nociones semánticas que la teoría naturalista de las representaciones lingüísticas de Stampe se había propuesto evitar. Esta es la evaluación y revisión de Fodor a la teoría de Stampe, revisión con la que estoy básicamente de acuerdo.

CAPITULO 6.

SOBRE CONTENIDOS Y REPRESENTACIONES MENTALES EN FODOR.

& 1. Hipótesis y límites de la investigación.

Según las secciones precedentes, los estados mentales que llamamos actitudes proposicionales son relaciones binarias entre un cierto sistema intencional y cierta clase de objetos que poseen una estructura lógica interna similar a la de las oraciones que en los lenguajes naturales ejercen la función de complementos de verbos que expresan actitudes proposicionales. Que las actitudes proposicionales tienen un objeto es lo mismo que decir que tienen un contenido. En virtud de este contenido, las actitudes proposicionales forman entramados de relaciones causales que, eventualmente, darán lugar a un comportamiento determinado y, en otras ocasiones, causarán nuevos estados mentales. Según Fodor, estos objetos son fórmulas de un lenguaje representacional interno. Estas fórmulas están constituidas por símbolos que tienen interpretación semántica y forma sintáctica, y lo que importa ahora es encontrar una teoría semántica naturalista, es decir, una teoría compatible con una metafísica realista y materialista, que dé cuenta del contenido de estas fórmulas. Por lo demás, decir que se puede naturalizar la semántica de estas fórmulas es lo mismo

que decir que se puede naturalizar la relación intencional, o, al menos, ciertas relaciones intencionales.

En el presente capítulo me propongo discutir con detalle la teoría naturalista que propone Fodor para explicar la intencionalidad y el significado de nuestros estados mentales, tal y como este filósofo la presenta y desarrolla de acuerdo con los planteamientos y con todas las restricciones que han aparecido hasta este momento. Por tanto, Fodor intenta solucionar con esta teoría el problema que con tanta sinceridad intelectual plantea y que caracteriza como crucial y definitivo para la fundamentación naturalista y fisicalista de las ciencias cognitivas.

La hipótesis de Fodor en el capítulo cuarto de su libro *Psychosemantics. The Problem of Meaning in the Philosophy of Mind* (1987) es que no hay ninguna razón *a priori* por la no podamos explicar la intencionalidad y el significado de los estados mentales en términos de relaciones causales. Este capítulo que he mencionado tiene un título muy representativo en este sentido: "El Significado y el Orden del Mundo". De acuerdo con Fodor, el hecho de que la muestra (token) de una palabra/símbolo en el Lenguaje del Pensamiento represente su objeto y tenga significado se debe a que se dan ciertas relaciones causales *relevantes* entre el símbolo mental y la cosa a la cual se refiere ese símbolo. Para circunscribir mejor el alcance de la investigación, Fodor centra especialmente su atención en el problema del significado de términos generales tales como "rojo", "caballo" o "gigante". Para usar la terminología de Fodor, el problema consiste, tal y como él lo ve, en buscar una teoría razonable acerca del vocabulario primitivo y no lógico de Mentalés (otra nombre que utiliza Fodor para hablar abreviadamente del hipotético Lenguaje del Pensamiento) en la que no se apele a nociones intencionales o semánticas.

En cierto sentido, la preocupación de Fodor es encontrar una teoría "atomística" del significado. En ningún momento ofrece una teoría acerca

de cómo podemos explicar en términos de relaciones causales el significado más complejo de las fórmulas derivadas. Por ejemplo, en las oraciones "x ama a y" e "y ama a x", el modo en que ponemos juntas 'x' e 'y' determina el significado de dichas oraciones. Y no sabemos cómo podemos explicar en términos causales lo que determina este modo de poner juntas las cosas. Fodor afirma, sin embargo, que éste no es ningún problema grave, y, desde luego, no ha de ser el primer problema. En el caso de los significados complejos lo que necesitamos es una semántica composicional, una semántica en la que las condiciones de verdad de los significados complejos venga dada en función de los significados de sus partes semánticamente relevantes. Para una semántica naturalista, el caso de los significados complejos de las fórmulas derivadas se resuelve cuando conocemos sus condiciones de verdad, y tenemos una teoría naturalista sobre el vocabulario primitivo y no lógico que aparece en estas fórmulas.

.... La intencionalidad de las actitudes se reduce al contenido de las representaciones mentales. Dada una definición de verdad, el contenido de las representaciones mentales está determinado por la interpretación de su vocabulario primitivo. De este modo, la interpretación del vocabulario primitivo, no lógico, de Mentalese es lo que está en el fondo del asunto, de acuerdo con la imagen presente. Correspondientemente, habríamos resuelto sobradamente el problema de la naturalización para una psicología de las actitudes proposicionales si fuéramos capaces de decir en un giro no intencional y no semántico qué es para un símbolo primitivo de Mentalese el tener una cierta interpretación en un cierto contexto. (Fodor 1987, 98).

& 2. Definición y división del problema.

La muestra de un símbolo mental denota su objeto porque hay una relación causal entre las muestras del símbolo y los objetos del mundo a los que se refiere dicho símbolo. Cuando un símbolo mental expresa una

instanciación de una propiedad en el mundo--digamos cuando "caballo" se refiere a la propiedad de ser un caballo, o "rojo" a la propiedad de ser rojo--la relación esencial es la adecuada relación causal entre el objeto individual en el mundo y el símbolo individual en la mente que lo expresa. En otras palabras, lo que necesitamos, de acuerdo con Fodor, es una relación contrafactual entre los símbolos mentales y los objetos del mundo, tal que (x) (y) (x representa y sii y causa x). Caracterizar esta relación legal es lo que se busca en primer término¹.

Antes de continuar quiero indicar que Fodor no está defendiendo la idea de que la noción de representación y la noción de causalidad sean meramente coextensivas. La coextensión no es una razón suficiente para la legalidad. Dos términos generales pueden ser coextensivos accidentalmente. Por ejemplo, el estar en un cierto lugar en un tiempo determinado puede ser coextensivo con tener un cierto color. Pero esta relación no implica que un objeto tenga un color determinado si y sólo si está en un cierto lugar en un cierto tiempo.

Para evitar posibles interpretaciones erróneas, conviene definir la cuestión que Fodor está intentando contestar con su teoría acerca de las representaciones mentales. La pregunta es ésta: ¿qué se requiere para que un sistema físico tenga estados intencionales del tipo de las actitudes proposicionales? (Fodor 1987, 97). Las actitudes proposicionales, como las

¹ Realmente, dicho así parece obligatorio abolir la arbitrariedad de la relación signo-denotación en el lenguaje. Desde luego, habría que explicar mucho más este punto, tarea que no es sino la especificación más fina del tipo de entidades de que consta el Lenguaje del Pensamiento. Sin embargo, explicar aquí este aspecto concreto de la teoría llevaría demasiado lejos, dado que su desarrollo conduce directamente a enfrentarse con otro más de los muchos problemas presentes en el conjunto de la teoría: el del innatismo de la semántica y de la sintaxis del Lenguaje del Pensamiento y su defensa. Para consultar los argumentos en favor del innatismo puede verse, entre otros, Jerry A. Fodor, 1975, The Language of Thought (Cambridge, Mass.: Harvard University Press): 55-97; 1987, Psychosemantics. The Problem of Meaning in the Philosophy of Mind (Cambridge Mass.: The MIT Press):135-154. También las teorías de Chomsky son, evidentemente, relevantes aquí. Véase, por ejemplo, Noam Chomsky, 1990, Language and Problems of Knowledge (Cambridge, Mass.: The MIT Press).

creencias, tienen condiciones de evaluación semántica. Ahora bien, de acuerdo con Fodor, una actitud proposicional es una relación entre un organismo y una fórmula del lenguaje interno del pensamiento. De esto se sigue que lo que prioritariamente se necesita conseguir son condiciones de evaluación semántica para estas fórmulas internas. Supuestamente, dichas fórmulas poseen una estructura lógica similar a la de las oraciones que en los lenguajes naturales cumplen la función de expresarlas. De esta manera, tenemos una guía para descubrir la estructura sintáctica de dichas fórmulas. El problema fundamental es especificar ahora las condiciones de su evaluación semántica de una manera tal que no haya necesidad de recurrir a nociones semánticas o intencionales, que es, de acuerdo con el programa naturalista, el objetivo que se persigue. La estrategia general es, por tanto, la siguiente: hay que establecer un contexto para las muestras de las fórmulas de Mentalés. Para establecer este contexto, es preciso especificar previamente una interpretación de las partes semánticamente relevantes que componen esas fórmulas. Así, la pregunta se resuelve en estos términos: ¿cuáles son las condiciones para tener una representación mental con un determinado contenido? La respuesta de Fodor es que una muestra de un símbolo mental tiene significado cuando es causada por, y covaría con, su objeto de una manera relevante. La explicación de esta afirmación y de su alcance para una teoría general de la intencionalidad, es el objeto de las páginas que siguen.

Para naturalizar el significado de nuestros estados mentales—brevemente, para naturalizar el significado mental (mental meaning)—lo que "necesitamos es una condición suficiente *plausible* para que 'A' exprese A; una condición tal que sea plausible que, al menos en algunos casos, los 'Aes' expresen A *porque se satisface esa condición*" (Fodor 1987, 111). Esta condición es, esencialmente, una cierta relación de covariación causal entre objetos naturales, y, en concreto, entre las representaciones mentales y sus

objetos. Cuando una instanciación de una cierta propiedad A, es causalmente responsable de un modo relevante de la aparición del símbolo 'A' en lo que Fodor llama la "caja de las creencias" (belief box), entonces se dice que el símbolo representa esa instanciación y que tiene, consiguientemente, determinado significado. Por supuesto, hablar de algo así como de la "caja de las creencias" en nuestro cerebro/mente es, hasta cierto punto, una broma. Fodor no está hablando literalmente de "cajas", sino sólo metafóricamente. Lo único que quiere dar a entender con esta expresión, o así lo creo, es que en virtud de ciertos procesos causales llegaremos a tener una relación de creencia con un cierto símbolo 'A', y que debido a esa relación el símbolo 'A' representará Aes en el mundo.

& 3. División de los conceptos.

Como Fodor señala, sería muy útil distinguir dos clases de conceptos o términos generales que deben ser explicados mediante una teoría causal de las representaciones mentales. La razón para establecer esta distinción es que la especificación de las condiciones causales suficientes para cada una de estas clases de conceptos es diferente. Fodor afirma que la primera clase de conceptos es aquella que encierra en sí todos aquellos términos que se refieren a objetos observables. Estos son los términos observacionales de Mentalés. Ejemplos de esta clase de conceptos son 'rojo', 'calor' o el concepto de la nota musical Do. En segundo lugar, están aquellos conceptos que no se refieren a una propiedad sensiblemente observable en el mundo, tales como, por ejemplo, los conceptos de 'protón' o de 'caballo'. En otras palabras, tenemos que distinguir entre *impresiones* e *ideas* en el más puro estilo empirista. Este modo de interpretar la teoría de Fodor dentro de la tradición empirista está justificado por el énfasis que pone este filósofo en las relaciones de covariación causal entre las

representaciones mentales y los objetos del mundo (Cummins 1989, 56).

De acuerdo con las teorías naturalistas del significado, el objetivo consiste en saber qué clase de relaciones causales relevantes deberían existir entre cada una de estas clases de conceptos y sus objetos.

& 4. Los conceptos observacionales.

En el caso de 'rojo', por ejemplo, Fodor conjetura que es nomológicamente necesario que una muestra del símbolo mental 'rojo' resulte o se dé en la mente, de grado o por fuerza, si es que ciertas circunstancias psicofísicas adecuadas son el caso. Tal y como lo utilizo en este contexto, el verbo "resultar" hay que entenderlo aquí simultáneamente en su doble significado de nacer, originarse o ser efecto una cosa de otra, y de aparecer o manifestarse una cosa. En este sentido y por lo que se refiere exclusivamente a esta clase de conceptos observacionales como, por ejemplo, 'rojo', Fodor sostiene que tenemos una ciencia, la psicofísica, que nos explica en un vocabulario no semántico y no intencional cuáles son estas circunstancias.

La psicofísica está precisamente interesada en decirnos cuánta <extensión> tiene que ser pintada de la pared ..., con qué intensidad de rojo tiene que ser pintada, y cuán próximos hay que estar a la pared y cuál tiene que ser el grado de luminosidad de las luces, y cosas así..., de tal modo que si se dan esa intensidad y esa luminosidad y uno está así de próximo ..., entonces pensará 'rojo' si sus ojos están dirigidos hacia la pared y su sistema visual está intacto. En una primera aproximación (las precauciones vendrán después), la psicofísica es la ciencia que nos dice cómo el contenido de la caja de creencias de un organismo varía con los valores de ciertos parámetros físicos de su entorno próximo. (Fodor 1987, 112-113).

En resumen, para naturalizar esta primera clase de conceptos que Fodor considera, necesitamos los siguientes requisitos:

(i) Cuando se dan cierto conjunto de circunstancias que son, o podrían llegar a ser, especificables por la ciencia psicofísica, entonces las instanciaciones de la propiedad ROJO "alojarán" (to token) muestras del símbolo mental 'rojo'. En este caso, el símbolo mental 'rojo' representará rojo.

(ii) Al menos en algunas ocasiones, dichas circunstancias se dan cuando casos individuales de la propiedad ROJO alojan muestras de 'rojo' en la mente.

(iii) Estas circunstancias son suficientes para explicar el hecho de que el símbolo 'rojo' represente rojo.

Si aceptamos esta función de la ciencia psicofísica dentro de una teoría naturalista general de la intencionalidad, entonces nos estamos acercando a pasos agigantados hacia una posible solución del problema planteado a propósito de la naturalización de las propiedades intencionales que exhiben las representaciones mentales. Fodor augura un futuro muy negro o, al menos, preocupante, para todos aquellos que piensan que hay razones metafísicas de peso para negar que puedan ofrecerse condiciones suficientes plausibles para explicar la intencionalidad de ciertos símbolos mentales, condiciones que son especificables sin apelar a nociones semánticas e intencionales (Fodor 1987, 113). En el vocabulario de la psicofísica sólo aparecen términos como "longitud de onda", "estado de la retina", "excitación neuronal", y otras por el estilo.

Alguien se podría echar las manos a la cabeza y razonar de la siguiente manera para protestar contra estas tesis. Es seguro que tenemos el concepto de rojo pero es muy probable que no tengamos todavía una ciencia psicofísica completa y verdadera. Y si esta situación se da, entonces, podría seguir preguntando este interlocutor, ¿qué tiene que ver con la

psicofísica que nuestras representaciones mentales representen lo que representan? A esto hay que decir lo siguiente. Es muy probable que la psicofísica esté todavía en mantillas, y es seguro que tenemos el concepto de rojo, pero este argumento no es un buen argumento contra la tesis de que "hay alguna clase de condición nomológicamente suficiente para que las instanciaciones de *rojo* causen muestras de 'rojo' que son psicológicamente activas" (Fodor 1987, 114). La teoría de Fodor sólo requiere que muestras del símbolo mental 'rojo' resulten en la mente en presencia de ejemplares de la propiedad ROJO cuando se dan ciertas circunstancias psicofísicas, y sostiene que la psicofísica puede especificar cuáles son estas circunstancias en un vocabulario no semántico y no intencional.

En el texto que he citado más arriba, el papel que se atribuye a la psicofísica consiste en explicar en un vocabulario no intencional las covariaciones que existen entre casos de rojo en el mundo y muestras del símbolo 'rojo' cuando estas muestras resultan en la *caja de las creencias*. Quiero subrayar estas palabras porque Fodor parece sugerir que 'rojo' representará rojo en caso de que aparezca una muestra de rojo en esa caja de las creencias. A primera vista, parece que importa poco, al menos para los efectos de una teoría naturalista del significado mental de los conceptos observacionales, que las muestras de 'rojo' se den en la caja de las creencias, de los deseos, o de las cosas que experimento de una manera pasiva. De acuerdo con las teorías representacionales de la mente, podemos tener diferentes actitudes proposicionales (creencias, deseos, etc.) hacia el mismo contenido.

De acuerdo. Sin embargo, Fodor afirma que la psicofísica, muy probablemente, no puede garantizar en todos los casos que alguien creará que "aquello es rojo" cuando mire a una pared suficientemente roja, suficientemente iluminada, etc. Pudiera suceder que, en contra de lo que las apariencias sugieren, no crea que "aquello es rojo". Pudiera ser que sepa

algo que le haga juzgar que aquello no es rojo al fin y al cabo (Fodor 1987, 114). Lo único que la psicofísica puede asegurar es que ese organismo experimentará *pasivamente* algo que expresaría diciendo cosas como "aquello es rojo"².

Pero una correlación con garantías entre instancias de rojo y muestras de "rojo allí" en la caja-de-lo-que-me-ocurre, sirve perfectamente bien para los propósitos de la naturalización semántica...Desde este punto de vista, la teoría de la fijación de las creencias estrictamente así llamada no pertenece a los semánticos, sino a los psicólogos cognitivos, uno de cuyos objetivos es decir qué determina cuáles de entre las cosas que me ocurren son las que de hecho vengo a creer (cómo la información fluye desde la caja-de-lo-que-me-ocurre a la caja de las creencias, si te gusta esta forma de decirlo). (Fodor 1987, 114).

En otras palabras, los mecanismos para fijar las creencias no son los mismos que los mecanismos psicofísicos. De cualquier manera, las muestras de 'rojo' en la caja-de-lo-que-me-ocurre representan rojo en la pared porque es imposible que alguien vea rojo y no lo vea *como* rojo en circunstancias

² Sería interesante relacionar adecuadamente la teoría de Fodor sobre las condiciones psicofísicas causales de las que depende que se origine y se manifieste la muestra del símbolo mental 'rojo' en el sistema de símbolos que constituiría el Lenguaje del Pensamiento, con la teoría de Husserl acerca de la receptividad en tanto que nivel ínfimo de la actividad del yo. Husserl, desde luego, no niega que, por ejemplo, la intensidad de un estímulo sea una condición necesaria para que se destaque el dato sensible y se produzca una tendencia afectiva del yo hacia ese dato. Es así como se destacaría éste del fondo homogéneo de la totalidad de datos sensibles del campo visual. El que un estímulo sea más o menos llamativo, más o menos nítido, esté a más o menos distancia de mí, etc. son factores relevantes para que esto suceda. Sin embargo, como Husserl dice en un lenguaje intencional, hay que distinguir entre "aquello que se impone y el yo al que se impone" (1985, 80). Husserl habla de que el estímulo puede atraer al yo con mayor o menor fuerza. El yo, por su parte, simplemente cede a los estímulos (1985, 81). En este sentido, la teoría de Husserl sobre las tendencias previas al *cogito* ("cuanto más intensa es esa "afección", tanto más fuerte es la tendencia hacia la entrega, hacia la "aprehensión" (Husserl 1985, 82)) y el sometimiento a la tendencia por parte del yo como estadio último en que vierte aquella tendencia, es, en el fondo, una idea muy semejante a la que expresa Fodor con su teoría. Hablar de "sometimiento", de "cesión" del yo, es otra forma de decir que se origina y aparece una muestra de 'rojo' en la caja de las creencias, de grado o por fuerza. Evidentemente, la diferencia que existe entre Fodor y Husserl es que Fodor pretende diluir y hacer desaparecer todo el vocabulario intencional del que se sirve Husserl para expresar sus descripciones y teorías.

psicofísicas óptimas. En circunstancias psicofísicas óptimas el genio maligno de Descartes no puede llevar a cabo su purificación epistemológica con la clase de conceptos que estamos analizando.

La cuestión que quiero suscitar es la siguiente. Aunque se den las circunstancias psicofísicas óptimas, y esas circunstancias aseguren las covariaciones entre 'rojo' en la caja-de-lo-que-me-ocurre y casos de rojo en el mundo, es todavía posible que las muestras de 'rojo' así caracterizadas no sean todavía conceptos que representen el rojo como una propiedad *de una pared roja*.

Hay condiciones óptimas que garantizan los contenidos en la caja-de-lo-que-me-ocurre. Estas muestras son "psicológicamente activas". Que sean psicológicamente activas quiere decir, o así lo entiendo yo, que cada vez que casos individuales de rojo en el mundo hacen que resulten, en condiciones psicofísicas óptimas, muestras del símbolo 'rojo', se da un *cambio* en los procesos mentales que computan la información sobre el mundo. Para aclarar más este punto podemos imaginar el siguiente experimento mental. Pensemos que pueda haber algo así como un observador "puro" del mundo, esto es, un observador sin ninguna mezcla de conocimiento teórico o preteórico sobre el mundo. Imaginemos que somos nosotros ese espectador "puro". Fodor piensa que en ese hipotético caso tendremos--y no podemos no tener--algunos conceptos bajo ciertas circunstancias acerca de las cuales la ciencia psicofísica puede dar cumplida cuenta (Fodor 1987, 117). Podemos imaginar la mente de un observador puro como una mente sin ninguna clase de conocimiento previo (background knowledge), una mente que sólo tiene la caja-de-lo-que-me-ocurre, por así decir. Esta idea cuadra bien con la tesis de que las relaciones causales psicofísicas entre rojo y 'rojo' no están mediadas teoréticamente. La mente de ese observador intacto puede tener la misma clase de muestras de representaciones mentales en la caja-de-lo-que-me-ocurre

que una mente con esas mismas muestras en la caja de las creencias. El problema, dicho crudamente, reside en saber si para una mente intacta las muestras de 'rojo' son símbolos/conceptos de las instanciaciones de la propiedad rojo en la pared.

Es interesantísimo constatar en este lugar que Husserl propone un experimento semejante en las *Investigaciones Lógicas* cuando trata el problema de las representaciones intuitivas (*anschauliche Vorstellungen*).

Si imaginamos una conciencia previa a toda experiencia, esta conciencia *siente* –en posibilidad– lo mismo que nosotros. Pero no intuye ningunas cosas ni acontecimientos objetivos; no percibe árboles y casas, ni el vuelo de los pájaros, ni el ladrido de los perros. Bien pronto nos sentimos impulsados a expresar la situación del siguiente modo: para una conciencia semejante, las sensaciones no *significan* nada; para ella las sensaciones no *valen como signos* de las propiedades de un objeto; para ella el complejo de sensaciones no funciona como signo del objeto mismo; las sensaciones son pura y simplemente vividas, pero carecen de toda *interpretación* (que la experiencia proporciona). (Husserl 1984 XIX/1, 80).

Si bien la similitud de los experimentos mentales de Fodor y Husserl es patente, sus respectivas conclusiones no pueden ser más dispares. La respuesta de Husserl a la cuestión suscitada es claramente negativa. La mera presencia o aparición de muestras de 'rojo' en la caja-de-lo-que-me-pasa no hace que dichas muestras representen el color rojo *de la pared*. En este caso, la simple presencia de 'rojo' no representa rojo, aunque dicha presencia modifique o pueda llegar a modificar el flujo de la información en nuestros procesos mentales y cause cambios en nuestros estados psicológicos en tanto que observadores intactos. Para Husserl, las cosas son, en este sentido, más complicadas. Puede haber muestras de 'rojo' presentes en la caja-de-lo-que-me-pasa que causen estados mentales funcionalmente distintos a los que causarían, digamos, muestras de 'púrpura' y, sin

embargo, es todavía posible que no tengamos el concepto 'rojo' o 'púrpura' alojado, sin que podamos hacer nada para ello, en la "caja" relevante.

Por ejemplo, podemos ser afectados por colores de los cuales no tenemos su concepto correspondiente, aunque los estemos viendo en circunstancias psicofísicas óptimas. En este caso, las muestras de estos colores resultarán alojadas en la caja-de-lo-que-me-pasa, pero no nos asistirán para que nos representemos el rojo como una propiedad de una pared roja. Pero esta última condición parece que es esencial para tener cierta clase de representaciones mentales, y las representaciones mentales es lo que nos incumbe aquí.

Otra dificultad paralela es que mucha gente llama 'rojo' tanto a casos particulares de paredes rojas, como a casos particulares de paredes escarlatas o púrpuras, aunque las instanciaciones de las propiedades de ser rojo, escarlata o púrpura deberían albergar muestras de símbolos diferentes en la caja-de-lo-que-me-ocurre y no la misma muestra 'rojo'. Este ejemplo parece sugerir que las muestras de 'rojo' en la caja-de-lo-que-me-pasa son, al menos, nociones muy complejas.

De acuerdo con estos ejemplos, ¿cuál es el papel de la psicofísica para las teorías semánticas naturalistas? Primeramente, la psicofísica sí que puede explicar en términos no intencionales y no semánticos los procesos computacionales que subyacen a la producción de las muestras de 'rojo' en la mente. Los estados de la retina, las excitaciones neuronales, las sinapsis, etc. se pueden considerar como elementos imprescindibles de los procesos computacionales que subyacen a la producción de nuestros conceptos observacionales. Es más o menos claro que a estos procesos computacionales no tenemos ningún acceso consciente directo. Pero, en segundo lugar, la función que Fodor atribuye a la ciencia psicofísica en esta clase de teorías naturalistas parece que no puede rendir los resultados que tan felizmente se proponía en un principio. A la luz de los ejemplos anteriores, la ciencia

psicofísica puede tropezar con dificultades a la hora de explicar la intencionalidad de símbolos/conceptos tales como 'rojo'. (Fodor habla efectivamente de conceptos y no de meras modificaciones funcionales en un sistema computacional.) La psicofísica puede especificar las condiciones óptimas que garantizan que las instanciaciones de la propiedad ROJO causarían covariaciones fiables en los estados funcionales de la mente/cerebro. Sin embargo, estas covariaciones puede que no sean suficientes para que muestras de 'rojo' en la caja-de-lo-que-me-pasa representen casos particulares de rojo.

Desde luego, no pretendo afirmar que lo que acabo de decir sirva como un argumento absolutamente concluyente en contra de la tesis de que la psicofísica es toda la teoría que necesitamos para explicar en términos naturalistas la intencionalidad de los conceptos observacionales. No creo que se pueda llegar a extraer esta conclusión tan fuerte de las reflexiones precedentes. Solamente digo que los ejemplos mencionados desafían al menos la plausibilidad del caso paradigmático que explota Fodor sobre el símbolo mental 'rojo'.

& 5. Los conceptos teóricos.

En contraposición a los conceptos observacionales, en el caso de los conceptos no observacionales no podemos justificar sólo mediante la existencia de circunstancias psicofísicas óptimas las cadenas causales que van desde los casos singulares, por ejemplo, de protones o caballos a los conceptos correspondientes de 'protón' o 'caballo'. El grado de complejidad y refinamiento de la teoría naturalista debe ser aquí mucho mayor.

La psicofísica puede garantizar las circunstancias en las que uno verá un caballo y en las que, presumiblemente, tendrá la apropiada experiencia caballuna (horsy experience). Sin embargo, no puede

garantizar el *contenido intencional* del estado mental en el que se está en aquellas circunstancias. Esto es debido a que no puede garantizar que cuando alguien ve un caballo, lo verá *como* un caballo. Dicho bruscamente, ver un caballo como un caballo requiere la aplicación del concepto CABALLO a lo que se ve. Y, por supuesto, los observadores intactos *qua* observadores intactos, *no tienen que tener* el concepto CABALLO. Así, pues, es perfectamente posible nomológicamente estar en una relación psicofísica óptima con un caballo y no tener todavía el pensamiento de que aquí hay un caballo" (Fodor 1987, 116).

Para explicar este caso voy a teorizar con las siguientes hipótesis como premisas.

(i) Las características intencionales de las muestras del símbolo mental 'rojo' se pueden explicar realmente de la manera que Fodor propone para naturalizar los significados mentales sobre observables.

(ii) No se puede explicar cierto conjunto de conceptos—posiblemente los más numerosos de toda nuestra vida mental—mediante procesos causales psicofísicos. Entre los conceptos que rehúsan esta clase de explicaciones podemos encontrar, por ejemplo, los de 'protón' y 'caballo'. Una de las características principales de los objetos a los que se refieren estos conceptos es que no son directamente observables. Objetos directamente observables son aquellos que pueden llegar a ser objeto de una experiencia sensible posible.

(iii) Tenemos conceptos de objetos que no son directamente observables y estos objetos existen en el mundo.

(iv) Es posible explicar el contenido de esta clase de conceptos en términos de relaciones causales sin apelar, por consiguiente, a nociones intencionales o semánticas.

La conclusión de estas hipótesis, en el caso de que fueran verdaderas, es que podemos ofrecer una teoría naturalizada del significado para casi todos los conceptos posibles. Todavía quedaría, muy posiblemente, un

importante conjunto de conceptos sin explicación: los conceptos matemáticos y lógicos.

Ahora la cuestión es saber cómo tiene que ser el mundo para que estas hipótesis sean verdaderas. Aquí está parte de la respuesta de Fodor a este problema. Hay una cadena causal fiable que conecta, primeramente, instancias de propiedades no observables--como, por ejemplo, instancias de PROTON--con instancias de propiedades observables--instancias de ROJO, por ejemplo-- y, en segundo lugar, otra cadena causal fiable que conecta estas propiedades observables con instancias del símbolo mental 'protón' que se refiere a aquella primera propiedad no observable.

Sin duda alguna, hay una falla en esta cadena causal y Fodor es plenamente consciente de ella. De acuerdo con este filósofo, lo que tendría que decir una teoría naturalista completa sobre estos casos sería que si A (protones en el mundo) causa de una manera fiable B (manchitas de color rojo en la cámara de niebla) y B causa fiablemente C (muestras del símbolo mental 'rojo'), y C causa fiablemente D (muestras del símbolo mental 'protón'), entonces A causa fiablemente D. Las covariaciones causales fiables son transitivas (Fodor 1987, 122).

Si aceptamos la hipótesis (i), entonces admitiremos también que la ciencia psicofísica puede dar cuenta de las cadenas causales fiables entre las instanciaciones en el mundo de las propiedades observables y su correspondiente concepto en la caja de las creencias. De esta manera, podemos explicar en términos no intencionales como es que 'rojo' se refiere a rojo. La psicofísica puede explicar, por consiguiente, cómo se puede pasar de B a C.

Sin embargo, todavía hay que andar algunos pasos más para explicar el cuadro completo. El siguiente paso es éste. Dado que es posible que nuestras teorías físicas sean verdaderas--o, al menos, que algunas partes importantes de nuestras teorías físicas lo sean--, entonces estas teorías nos

dicen, básicamente, como es que A causa fiablemente B, como es que los protones causan muestras de rojo en la cámara de niebla. Ahora la tarea consiste en saber cómo podemos explicar que C--muestras del símbolo mental 'rojo'--cause fiablemente D--muestras del símbolo mental 'protón'. Si conseguimos este objetivo, entonces podemos explicar por qué 'protón' se refiere a protón. Una vez más, este fin hay que poder alcanzarlo sin recurrir a ninguna propiedad semántica o intencional. Tiene que haber, por consiguiente, un sendero causal fiable entre las muestras de 'rojo' y las muestras de 'protón'.

Según Fodor, el eslabón perdido es, precisamente, una teoría física.

Lo que necesitamos es un algún proceso que sitúe una muestra de 'protón' dentro de la caja de creencias siempre que un protón en el mundo sea causalmente responsable (via leyes físicas y psicofísicas) de que haya una muestra de 'rojo' en la caja de creencias. Ahora bien, lo que normalmente lleva a cabo esta función es, de hecho, una inferencia basada teóricamente: el observador sostiene una teoría que implica que las muescas de rojo señalan protones y él aplica lo que sabe para inferir la presencia del protón a partir del color de la muesca. Cuando la teoría es verdadera, las inferencias que median tendrán un éxito real para establecer correlaciones entre las instanciaciones de *protón* en el entorno experimental con las muestras de 'protón' en la caja de las creencias. Las teorías verdaderas--cuando se internalizan--establecen correlaciones entre el estado de la cabeza y el estado del mundo: Esto es precisamente lo que ellas tienen de estupendo. (Fodor 1987, 121).

De acuerdo. Pero "teoría", "inferencia", "mediación" y "conocimiento" son nociones semánticas. Esta solución, como Fodor señala, es la que una teoría naturalizada sobre el significado no puede permitir. Es también la clase de solución que Cummins piensa que es circular y desastrosa para las teorías naturalistas de las representaciones mentales. El razonamiento de Cummins es como sigue. Fodor dice que hay representación cuando hay covariación causal entre el símbolo mental y su objeto, entre representación

y *representandum*. En el caso de los conceptos no observacionales, la covariación depende de un mecanismo que, bajo condiciones de buen funcionamiento, hará que resulte una muestra de 'protón' a partir de un caso de protón. Dado que este mecanismo es un mecanismo de inferencia fruto de la internalización de una teoría, no puede haber inferencia si no hay ya representaciones internas gracias a las cuales se pueda llevar a cabo estas inferencias. La consecuencia de este planteamiento es que para entender los mecanismos que explicarían la covariación causal entre 'protón' y protón debemos entender ya lo que es una representación y el papel que desempeña en los procesos mentales. Y esto, según Cummins, es circular (Cummins 1989, 65).

¿Qué salida queda? Claramente todo el problema se centra en el papel mediador de las teorías en esta clase de consideraciones. Las teorías científicas o las pseudoteorías de la vida cotidiana pueden ser verdaderas o falsas. Si la teoría no fuese verdadera, lo que sucedería es que las muestras de 'rojo' en la caja de las creencias, aunque sean causadas fiablemente por los casos de rojo que aparecen en la placa fotográfica, y aunque dieran lugar, mediante ciertos mecanismos, a la aparición del símbolo mental 'protón', este símbolo no se referiría a los protones sino, por ejemplo, a tomates microscópicos. Veamos. Algunas personas seguramente defienden teorías muy divertidas y un poco locas, por no decir completamente falsas, que les pueden llevar a pensar que cada vez que ven una muesca de color rojo en la placa fotográfica, estas muescas son causadas por microexplosiones de tomates microscópicos tan pequeños como los protones. Sin embargo, es muy probable que las teorías falsas tengan también sus axiomas, teoremas y reglas de inferencia. En los casos en los que la relación entre un concepto y su objeto esté mediada teóricamente, tendremos que explicar cómo se puede naturalizar esta mediación teórica, tanto si su resultado es

verdadero como si es falso. Esta es la solución de Fodor a esta clase de dificultades.

La afirmación, para ponerla de una manera brusca pero relativamente intuitiva, es que para que 'protón' exprese *protón* es suficiente con que haya una correlación fiable (reliable correlation) entre protones y 'protones' que sea producida por un mecanismo cuya reacción sea específica a las huellas psicofísicas de las cuales los protones son, *de hecho*, causalmente responsables. Y *esa* afirmación *puede* hacerse en un vocabulario no semántico, ni intencional....

Sin duda, los mecanismos que rastrean o siguen la pista a los conceptos de objetos no observables de la forma requerida satisfacen caracterizaciones intencionales (son típicamente inferenciales) y caracterizaciones semánticas (funcionan porque las inferencias que extraen son buenas). Pero eso está bien porque, por una parte, las propiedades semánticas/intencionales de tales mecanismos son, por así decir, condiciones *contingentes para que tengan éxito rastreando protones* y, por otra, porque lo que se requiere para que 'protón' exprese *protón* es sólo que el rastreo sea, de hecho, exitoso. Para los propósitos de la naturalización de la semántica, *es la existencia de una correlación mente/mundo fiable lo que cuenta y no los mecanismos por los que esa correlación se produce.* (1987, 121-122).

Los mecanismos a los que alude Fodor son los ordenadores. Podemos naturalizar los significados diciendo que las teorías son máquinas para calcular. Estas máquinas calculadoras se pueden considerar como sistemas asociacionistas, Y "asociación" no es una noción semántica. La idea básica de Fodor es que la función que cumplen las teorías, cuando se internalizan, a la hora de establecer "el contenido semántico de los conceptos es tan meramente mecánica que incluso hasta un asociacionista podría reconstruirla" (Fodor 1987, 165).

El problema, entonces, es éste. Asumamos que hay ciertos mecanismos que son responsables de que se dé como respuesta 'protón' cuando su "input" es 'rojo'. Estos mecanismos pueden ser perfectamente asociaciones no semánticas entre enunciados, como sugiere Fodor siguiendo

a Quine. El papel de las máquinas asociacionistas es, precisamente, transferir un cierto estado de activación que parta de los "inputs" y de los conocimientos previos con los que cuenta el sistema, y que median también en el proceso, y que haga que el resultado final de la computación sea 'protón'. Los mecanismos asociacionistas son procesos computacionales, esto es, están regidos por reglas que manipulan símbolos en su aspecto meramente formal, sintáctico. Una muestra del símbolo mental 'protón' representa protón, porque una muestra del símbolo mental 'rojo' causa, de una manera asociacionista no semántica, una muestra del símbolo mental 'protón' cuando un protón causa, de hecho, una muesca de color rojo en la placa fotográfica.

Todo esto está muy bien, pero todavía falta, en mi opinión, demostrar, o dar al menos, argumentos convincentes que hagan digerible la idea de que esos mecanismos computacionales conectan *de una manera fiable* 'rojo' con 'protón'. No es suficiente decir que ciertos procesos computacionales alojan 'protón' en la caja de las creencias cuando su "input" es 'rojo'. La razón de esta insuficiencia está en que estos procesos pueden seguir senderos determinados por teorías completamente locas—recuérdese el ejemplo de los tomates microscópicos—, o seguir senderos causales no usuales³. La consecuencia en ambos casos sería que los símbolos mentales para los que estamos tratando de establecer un contexto no intencional no representarían sus objetos. Los procesos tienen que correlacionar los símbolos mentales con sus objetos de una forma fiable. En este sentido, podría suceder o, al menos, sería pensable, que las leyes de acuerdo con las cuales funcionan los procesos formales asociacionistas y las leyes que gobiernan las inferencias fiables de las teorías físicas

³ Berenc Enc ha examinado recientemente el problema que se plantea a la hora de especificar el sendero causal correcto. Toda teoría causal de las representaciones tiene que enfrentarse tarde o temprano con este problema. Véase su "Causal Theories and Unusual Causal Pathways." *Philosophical Studies* 55 (1989): 231-261.

supuestamente verdaderas, sean completamente dispares. Necesitamos, por consiguiente, una *prueba* de la fiabilidad si es que queremos suscribirnos a la teoría de Fodor. En este punto, los problemas y compromisos son los mismos, o muy similares, a los que tienen que enfrentarse las teorías instrumentalistas en matemáticas *à la Hilbert* (véase Detlefsen (1986)). Los mecanismos asociativos son, como ya dije, máquinas que procesan símbolos puramente formales. Pero estos mecanismos tienen que asegurar que sus "outputs" tienen además valor epistemológico. Si tienen valor epistemológico, entonces también tendrían valor representacional. Necesitamos, como mínimo, una prueba instrumental de la fiabilidad de estos mecanismos. Si conseguimos esta prueba, podremos explicar las condiciones de producción de las muestras de 'protón' haciendo uso de un vocabulario que no sea ni intencional, ni semántico. Esto es lo que, a mi juicio, le falta a la interesantísima teoría de Fodor y es mi diagnóstico final. Pero es algo que, por desgracia, no sé cómo realizar y curar. Como otras tantas veces en filosofía, es más fácil detectar los posibles problemas que solucionarlos.

CUARTA PARTE.

EL PROBLEMA DE LA CONCIENCIA.

CAPÍTULO 7.

DOS TEORIAS DE LA CONCIENCIA: DENNETT Y SEARLE.

1. LA TEORIA DE DENNETT.

"Il n'y a rien de si merveilleux que le Mécanisme de la Nature ne soit capable de produire" (Leibniz 1960, 204)

& 1. Psicólogos y filósofos.

Es difícil saber en qué punto o puntos concretos se necesita introducir el problema de la conciencia cuando se trata acerca de las cuestiones que genera la investigación de la intencionalidad de los estados mentales, sobre todo cuando tales cuestiones se plantean de la manera en que aquí se ha hecho. Desde aquellos filósofos y psicólogos que tratan de evitarlo con mayor o menor disimulo—y con mayor o menor fortuna—, hasta aquellos otros que se encaran directamente con él, se tiene la vaga idea de que, en cualquier caso, la conciencia es algo que sólo viene a fastidiar la fiesta. Por supuesto, incluso entre los decididos, hay una gran disparidad de opiniones acerca de qué es la conciencia y qué tiene que ver con todo este berenjenal. He aquí a uno de los decididos, para quien el problema tiene, al menos, sentido: Daniel C. Dennett.

El cognitivista debe tomar en serio a la conciencia, pero hay maneras relativamente poco comprometedoras de hacerlo. (1981, 150).

Según el diagnóstico de Dennett y hasta donde se sabe, hay un abismo, de momento difícil de salvar sin quebraderos de cabeza, entre las investigaciones psicológicas empíricas que están orientadas hacia la construcción de los mejores modelos teóricos funcionales que den cuenta de la multiplicidad de los datos empíricos disponibles, y la conciencia de los sujetos de experimentación. Según se comentó con cierto detalle en el capítulo dedicado a exponer algunas cuestiones metodológicas importantes que gravitan en torno a la psicología como ciencia y, en concreto, en torno a la psicología cognitiva como la "nueva" ciencia psicológica, a los sujetos de experimentación no se les pedía más que atendiesen a uno de los dos auriculares por los que se emitía un mensaje, que indicaran cuándo veían más grande o más pequeña una mancha, o que señalaran con el dedo dónde veían ellos un objeto determinado. Como se recordará, la realización de estas tareas por parte de los sujetos tenía lugar en condiciones heterofenomenológicas peculiares, en circunstancias siempre determinadas por el hecho de que el psicólogo que ha diseñado el experimento tiene un conocimiento más detallado y amplio de las condiciones y variables a las que está sometido el sujeto. El psicólogo posee, pues, un conocimiento más amplio y detallado de cómo es el mundo que el que tiene el sujeto de experimentación. Desde este punto de vista, la "reflexión" del sujeto de experimentación sobre las propias vivencias intencionales y la posterior expresión de sus contenidos conscientes eran simplemente razones adecuadas que daban pie al psicólogo para postular cuáles eran los procesos psicológicos que debían de mediar la aparición de lo que los sujetos dicen y hacen--de su experiencia consciente y de su conducta intencional en definitiva--, así como cuáles eran las propiedades más relevantes de estos

procesos (mentales)¹. El investigador debía proceder así si es que quería explicar, utilizando genéricamente el método hipotético-deductivo, las correlaciones halladas entre las condiciones experimentales y las respuestas de los sujetos en esas condiciones.

Esta manera de proceder indica claramente que los psicólogos de orientación cognitiva no sienten ningún rubor especial a la hora de utilizar *metodológicamente* los enunciados introspectivos y las acciones intencionales de todo tipo como una fuente de datos valiosa, si no esencial, para elaborar sus teorías. En cualquier caso, y tal y como se puede apreciar, la conciencia, dentro de esta metodología, no constituye un objeto de estudio cerrado y relativamente independiente al que fuese posible dedicarle una ciencia igualmente independiente y los esfuerzos conjuntos de una comunidad científica que investigase sistemáticamente las propiedades de los objetos que caen dentro de dicho dominio, cualquiera que éste sea. La conciencia es más bien, repito, una fuente de datos muy valiosa--pero una entre varias otras también muy importantes--para hacer otra cosa diferente: psicología cognitiva, ciencia cognitiva. Es de esta manera como los psicólogos, sin dejar de presentar sus protocolos respetuosos a la conciencia y de hacerle los honores debidos, han intentado posteriormente eludir deliberada y sutilmente los problemas que ella pueda plantearles en el desarrollo efectivo de sus intereses teóricos y empíricos (Dennett 1981, 149).

Semejante actitud ha propiciado una división del trabajo en la sociedad académica. Junto a los psicólogos y a los científicos cognitivos en general se supone que están los filósofos--desde luego, no todos ellos--quienes se ocuparían de no dar la espalda a la conciencia y de mantenerse fieles a ella de palabra y obra. Ella no es medio para nada, sino fin en

¹ En contra de esta idea, Colin McGinn ha argumentado que "la propiedad de la conciencia está cognitivamente cerrada (cognitively closed) con respecto a la introducción de conceptos que tienen a su base inferencias extraídas de la mejor explicación" de los datos (1991, 13). Sobre la noción de cierre cognitivo (cognitive closure) véase (McGinn 1991, 3).

sí misma. Incluso, de hacer caso a los filósofos de orientación fenomenológica, ella es el único objeto propio de todas las investigaciones filosóficas rigurosas dignas de este nombre. Sin embargo, es con esta división del trabajo con la que no se siente a gusto Dennett. Las divisiones sólo pueden causar rencillas y malos entendidos y, al final, generan apatía, encono y desinterés por aquello por lo que parece desvivirse el partido opuesto. Esta es la situación que Dennett quiere evitar a toda costa.

Según este filósofo, una mirada retrospectiva a la Historia de la Filosofía de este siglo hace posible deslindar, al menos, tres actitudes que los filósofos han adoptado en relación a las formas posibles de abordar los problemas de la filosofía de la mente; tres posiciones ante las cuales tiene que tomar partido cualquier filósofo interesado en el problema de la intencionalidad y de la conciencia. Estas actitudes quedan así claramente expuestas:

El filósofo de la mente tenía tres elecciones: abandonar la filosofía y perseguir teorías empíricas en el dominio de la psicología o de la ciencia del cerebro, abandonar la teoría y fijarse en las modestas iluminaciones y curas de confusiones de los análisis puramente lingüísticos, o llegar a ser una especie de metateórico, un crítico conceptual de las teorías empíricas propuestas por las ciencias relevantes. Es esta última concepción de la empresa, que es vista como una rama de la filosofía de la ciencia, la que domina hoy el mejor trabajo en este campo. (Dennett 1978, 250).

Por mi parte, añadiría una cuarta opción que casi siempre se descarta cuando se discuten estos temas, si es que alguna vez llega a ser tomada realmente en cuenta, a parte de su mención puramente nominal. Me refiero a la posibilidad de seguir haciendo fenomenología según el camino emprendido definitivamente por la filosofía husserliana, sin por ello caer en una actitud acrítica respecto de las investigaciones que le son propias y sin cancelar la necesidad de ampliarlas con nuevas fuentes de problemas y

datos. No hay por qué descartar que la última filosofía de la psicología, la última iluminación de las cuestiones de principio de la psicología como ciencia objetiva no haya que ir a buscarla y, eventualmente, encontrarla en este otro ámbito de estudio, por mucho que, por otra parte, haya que ampliarlo e, incluso, reformarlo. Digo esto, porque no conviene dejar ninguna alternativa fuera del tinero cuando uno se toma la molestia de representarse las opciones disponibles en relación con este asunto, máxime cuando la elección final puede influir de una manera tan considerable, por no decir decisiva, en la clase de filosofía que cada uno, en tanto que filósofo o simplemente pensador, vaya a hacer en el futuro; y, por consiguiente, en la caracterización de aquello a lo que va a dedicar el tiempo que reserva a la comprensión del mundo y de sí mismo.

Además de estos aspectos más personales, la elección es, en cualquier caso, una elección que influye directamente en la consideración de cuál es el método general de la filosofía, si es que hay que asignarle propiamente alguno. Al menos, por lo que respecta a esta última cuestión, la respuesta de Dennett en relación al ámbito de los problemas propios de la intencionalidad no puede ser más clara. A juzgar por los resultados obtenidos hasta la fecha, el mejor método que puede emplear la filosofía para estudiar la intencionalidad es una investigación de carácter metacientífico que siga de cerca, pero críticamente, los métodos y resultados de la investigación teórica y empírica en el dominio de la psicología y de las demás ciencias cognitivas. En este círculo de investigaciones queda, desde luego, bastante por hacer, aunque se ha hecho ya mucho. La adopción definitiva de este método filosófico, el interés mostrado abiertamente por los filósofos de la mente hacia las teorías y multiplicidad de datos empíricos que aportan la psicología cognitiva, la neurofisiología, la inteligencia artificial, la lingüística y, en general, el conjunto de las denominadas "ciencias cognitivas", ha generado la diferencia fundamental que

ve Dennett entre los trabajos más importantes que se han escrito siguiendo esta orientación filosófica y los elaborados desde la perspectiva de las otras dos, o, según mi parecer, tres otras posibles opciones.

Pero no es solamente respecto del problema de elaborar una teoría de la intencionalidad en general, sino, en especial, dentro de ésta, respecto del problema de elaborar una teoría de la conciencia, donde ve Dennett la posibilidad de utilizar con provecho la reflexión metateórica sobre los resultados de la psicología cognitiva. Detallar esta posibilidad y orientar cómo habría de desarrollarse en concreto, es el fin que se propone este autor en uno de los trabajos más importantes, detallados e influyentes que se han publicado hasta la fecha sobre el tema: "Hacia una teoría cognitiva de la conciencia". De una manera singular, el puente de unión que permitiría superar el abismo que atisba Dennett entre la investigación cognitiva concreta y las peculiaridades de la conciencia sólo lo puede trazar y construir el filósofo/psicólogo que adopta de una vez por todas la actitud metateórica.

¿Cómo plantea Dennett el problema de la conciencia para que éste logre ser accesible y tratable con las herramientas de análisis que provee la adopción de un punto de vista metateórico en filosofía de la psicología? ¿Cómo este método puede conducir finalmente a la elaboración de una teoría de la conciencia que sea "relativamente poco comprometedora"? Pero, ¿comprometedora, respecto de qué? ¿Con qué no quiere comprometerse Dennett *demasiado*? El lo dice claramente. Por nada del mundo le gustaría comprometerse con una concepción de la conciencia en la que no se permitiera su estudio empírico, con una aproximación a la investigación racional de la conciencia en la que se diese la espalda a la ciencia, a sus métodos y resultados, por muy provisionales que éstos sean (Dennett 1982b, 159).

& 2. Tres clases de acceso a un contenido.

Yo soy consciente de esto o de aquello. Esta breve fórmula resume el problema entero de la conciencia. Si yo fuera el sujeto de alguno de los experimentos psicológicos que se analizaron en el capítulo segundo, podría decir, en contestación a las tareas propuestas, que yo soy consciente de que entiendo lo que se me dice por uno de los auriculares, que yo soy consciente de que la mancha es ahora más grande y ahora más pequeña, que yo no puedo evitar señalar con el dedo a donde señalo, etc. Dennett escribe de nuevo la escueta expresión "yo soy consciente de esto o de aquello" como "yo soy consciente de aquello a lo que yo tengo *acceso*" (1981, 150)—subrayando especialmente el *yo*— y procede a caracterizar qué quiere decir este acceso del yo a sus contenidos conscientes. Y todavía más que investigar qué quiere decir este acceso, se trata de averiguar qué es él realmente y no meramente cómo estaríamos dispuestos u obligados a describirlo.

Según Dennett hay tres formas posibles de tener acceso a un contenido o información. El acceso de la conciencia personal, el acceso computacional y el acceso público. Si se deja temporalmente de lado la descripción y explicación del acceso consciente, que ocupará de por sí suficientes páginas en breve, y nos centramos en las otras dos clases de acceso, que son las que más peso han tenido en la elaboración concreta de las diversas teorías cognitivas, la caracterización de estas dos últimas especies de acceso queda así expresada. Se dice que una de las subrutinas del programa total tiene acceso computacional a los resultados o información de otra u otras subrutinas cuando los "utiliza" para continuar o, eventualmente, terminar la computación. Si se concede que los procesos que suceden en el sistema nervioso, o en cualquier otra clase de sistemas materiales complejos a los que se adscribe contenido, admiten una

descripción funcional abstracta adecuada como procesos computacionales, entonces "el acceso computacional no tiene nada que ver *directamente* con el acceso de la conciencia personal, ya que *nosotros* no tenemos acceso a muchas cosas a las que varias partes de nuestro sistema nervioso han mostrado tener acceso" (Dennett 1981, 151). A simple vista, por lo tanto, no es de gran provecho servirse del acceso computacional para dar cuenta del acceso peculiarísimo que la conciencia tiene a sus propios contenidos. De ser ésta la situación de hecho, habría que buscar en otra dirección.

El acceso público, por su parte, es aquel que el usuario o el programador tiene a los estados computacionales transicionales por los que pasa el programa con el fin de saber lo que el ordenador está haciendo en cada momento. Esto se consigue gracias a la programación de una nueva subrutina que tiene acceso computacional, más o menos parcial, a aquellos estados computacionales transitorios sobre los que se quiere tener información. Todo lo que hace falta es que la información que se obtiene gracias a esta subrutina se pueda imprimir para su lectura y uso público posterior. La clase de metáfora política de que se sirve Dennett para distinguir entre el acceso computacional y el acceso público a la información--aplicada al caso español--hace que pensemos en Felipe González como alguien que tiene acceso computacional a una cantidad enorme e importantísima de información de la que nosotros, el público, sólo tenemos un conocimiento más o menos amplio--aunque más bien pequeño--gracias a la información a la que Rosa Conde, la Ministra Portavoz del Gobierno, tiene acceso y quiere transmitirnos, o a la que podemos encontrar impresa a diario en el B.O.E. Respecto de la noción de acceso público en relación a la conciencia personal y a sus contenidos, comenta Dennett lo siguiente:

La noción de acceso público parece acercarnos más al acceso personal de la conciencia, porque somos criaturas que hablan (tenemos una especie de facultad de impresión), y--al menos en una primera

aproximación--aquello de lo que somos conscientes es aquello de lo que podemos hablar, introspectiva o retrospectivamente. (1981, 152).

Con todo, no es del acceso computacional de una subrutina a la información de otra subrutina, ni del acceso público, tal y como éste se acaba de caracterizar, de lo que parece que hablamos cuando queremos referirnos al acceso privilegiado que el "ojo interior" de la conciencia tiene a sus propios contenidos cuando "se vuelve" sobre ellos. El sujeto, el yo que tan íntimamente está ligado a la conciencia de poseer determinados contenidos, no puede identificarse--como muy bien apunta Dennett--ni con el público que tiene acceso indirecto a la información que fluye dentro de un sistema gracias a su acceso directo a los medios que divulgan esta información, ni con la facultad de impresión que tiene acceso computacional a la información sobre el estado temporal del sistema (1981, 152). Ni creemos, ni queremos que el yo, el yo que somos cada uno de nosotros--por lo que se refiere exclusivamente a su vertiente cognitiva y a los procesos psicológicos *más básicos*--, se limite simplemente a ser, tras un análisis racional, una compleja malla organizada de partes o subrutinas entre las que hay un intercambio de información, sea como sea que haya que describir respectivamente esta malla y esta información. La organización de un Ministerio, cuando éste funciona como un todo, se podría adecuar a esta descripción en un sentido relevante y, sin embargo, nadie estaría dispuesto a otorgarle conciencia personal. Sin embargo, en un principio, nada hay tampoco que impidiera describir esta malla, para el caso de los diferentes subsistemas que están implicados en el funcionamiento del todo que es el organismo psicofísico, con el detalle suficiente que exige el fin de la construcción de los modelos teóricos psicológicos funcionales que tienen precisamente como objeto de análisis este organismo cuando se le considera como organismo psíquico-físico y no meramente como sistema físico.

Desafortunadamente, la dificultad suprema que hay que encarar de nuevo cuando se procede así es que tampoco en este caso--al igual que en el caso del Ministerio--estamos dispuestos a atribuir ninguna clase de conciencia personal al sistema funcional que resultara de la coordinación de los diferentes subsistemas sobre los que disponemos actualmente de teorías psicológicas y cognitivas plausibles.

Ciertamente, Dennett se compromete sin reservas con uno de los seis dogmas fundamentales del ser psíquico que se destacaron en el capítulo segundo. El funcionalismo le parece una tesis irrenunciable en filosofía de la mente, la única tesis suficientemente refinada que permitiría seguir sosteniendo una ontología respetablemente materialista sin traicionar las peculiaridades y anomalías que introduce el ser psíquico en el orden causal de la naturaleza. El problema con este modo de proceder es, sin embargo, que todas las teorías funcionalistas son, sin excepción alguna, teorías "subpersonales", como Dennett las denomina. Y, en cuanto teorías "subpersonales", sólo emplean dos de las nociones de acceso que se detallaron más arriba, la de acceso computacional y la de acceso público. Así, pues, que no extrañe a nadie pensar que si las teorías cognitivas sobre el ser psíquico implican, cuando se reflexiona metateóricamente sobre sus métodos y compromisos ontológicos, la tesis del funcionalismo, se tenga la sospecha bien fundamentada de que ninguna teoría psicológica valdría para explicar el acceso de la conciencia personal a sus contenidos. Dennett introduce además la reseva de que, aun cuando haya teorías cognitivistas que no se comprometen con el funcionalismo, sin embargo, toda teoría cognitivista, hasta donde él sabe, es una teoría del nivel subpersonal (1981, 153). De permanecer inamovible la plausibilidad de la tesis que defiende que no hay teoría cognitivista que no se desarrolle enteramente bajo el supuesto de que todo lo que ella puede hacer se mantendrá dentro de los severos límites impuestos por los análisis subpersonales computacionales y públicos,

entonces queda siempre ilesa la intuición de que ninguna ejemplificación posible de un modelo teórico psicológico idealmente terminado y realizado en un sistema físico suficientemente complejo, dará como resultado un yo personal, "una vida consciente e interior" (Dennett 1981, 154). Si se acepta este resultado, el problema del acceso del yo a los contenidos propios de los que él es consciente parece que cae de nuevo fuera del ámbito de intereses de la psicología científica y, desde luego, más allá de lo que ésta puede hacer con los recursos conceptuales de los que dispone hasta la fecha. Claramente, con este punto de partida, la adopción posterior de una perspectiva filosófica metateórica en relación a la ciencia cognitiva no podría rendir nunca el puente deseado entre la teoría y la conciencia. Es así como la división del trabajo queda otra vez restablecida. Los psicólogos cumplirían su trabajo y los filósofos que no piensen que todo lo que les resta por hacer en relación al problema de la intencionalidad es seguir la evolución de las teorías psicológicas y desentrañar sus compromisos ontológicos, harían el suyo, que no es otro que ocuparse de la conciencia. Dennett admite que este resultado es, en cierto modo, inapelable.

Esto es verdad: la "teoría" del nivel personal para las personas no es una teoría psicológica. (1981, 154).

No solamente este resultado sería inapelable, sino que, a primera vista, lo sería por principio. Sin embargo, es precisamente este resultado el que debe ser capaz de sortear ariosamente Dennett si es que quiere seguir manteniendo que la división del trabajo a que daría lugar un estado de cosas semejante no es en absoluto deseable ni en el panorama intelectual de un psicólogo, ni en el panorama intelectual de un filósofo. Su objetivo, por consiguiente, no es otro que el de "construir un "yo" hecho y derecho a partir de partes subpersonales, explotando las nociones subpersonales de

acceso ya introducidas <el acceso computacional y el acceso público>" (1981, 154). ¿Cómo se realizaría esta tarea?

& 3. Un mundo subpersonal para la conciencia.

Reconozco que el planteamiento del problema por parte de Dennett es sorprendente. Reconozco incluso que también me entusiasma por su penetración, por las claras divisiones que introduce y por cómo prepara el camino para un tratamiento despersonalizado o "subpersonalizado" de la conciencia. Para seguir avanzando por esta senda, la estrategia que sigue Dennett no es otra que la de diseñar un diagrama de flujo, un modelo simplificado de los procesos cognitivos sobre los que se dispone de alguna teoría suficientemente elaborada, cuya descripción se realiza en un nivel enteramente subpersonal, pero que sirva posteriormente para hacer indicaciones precisas acerca de cómo podría tratarse el problema de la conciencia personal de una forma relativamente poco comprometedora. Característico de este diagrama de flujo es que en él no hay nada que represente a este yo que todos estamos deseando que nos digan qué es, si es que es algo. Supongo que filósofos como Wittgenstein han mostrado suficientemente que cualquier dibujo o diagrama que sitúe en algún lugar del papel al sujeto es una representación completamente inadecuada del mismo (Tractatus 5.6331). Dennett, desde luego, ha aprendido bien la lección. En su diagrama no aparece ningún puntito, ni ojo, ni cosa parecida, que represente al yo. Sin embargo, sigue pensando que se puede construir un "yo" como Dios manda simplemente con seguir las funciones y las interrelaciones entre las diversas partes que constituyen este diagrama de flujo. La posibilidad de realizar esta construcción hace que se pueda hablar de la psicología como ciencia del yo. Si es que puede haber teoría del yo, entonces llamaré a esta ciencia "egología". La psicología científica es egología empírica y, de

creer a Dennett y a Rorty, en verdad la única egología posible y deseable.

Los detalles en este caso no son, en principio, demasiado importantes. Cójase un buen manual de psicología cognitiva al uso, sígase con cuidado las teorías más importantes que se proponen en cada uno de sus capítulos, póngase en relación mutua en forma de organigrama cada una de estas teorías y dibújese, con un poco de imaginación, el resultado final. Si se procede así se habrá obtenido seguramente una variante de la estructura funcional global que propone Dennett para el psiquismo humano de acuerdo con los modelos y teorías con los que cuenta hasta hoy la psicología y las ciencias cognitivas (1981, 155). En el organigrama, por ejemplo, habrá una casilla para representar "el lugar" de la memoria a corto plazo; es decir, el "lugar" donde se lleva a cabo una primera organización precategorial, de capacidad limitada y de escasa duración, de la información sensorial que procede del entorno inmediato en que vive el organismo. Se hablará a este respecto, quizás, de memoria icónica y ecoica según la modalidad sensorial. Esta clase de memoria a corto plazo se distinguirá en varios puntos importantes de la memoria a largo plazo, a la que también se le asignará otra casilla en el organigrama general. Si la memoria a corto plazo se caracteriza por su duración y capacidad limitadas, la memoria a largo plazo, por el contrario, tiene una capacidad de almacenaje muy superior y la información que se alberga en ella puede ser recuperada, aunque pase mucho más tiempo. Para dar cuenta de estas diferencias entre la memoria a corto y a largo plazo, muchos modelos cognitivos postulan que la información almacenada en una y otra clase de memoria ha de ser codificada utilizando estructuras de datos muy diferentes. Esta diversidad en la estructuración permitiría que la información albergada en la memoria a largo plazo permanezca en la mayoría de los casos en un estado inactivo y que algunas de sus partes puedan ser recuperadas cuando las circunstancias ambientales en las que se encuentra el organismo así lo

requieran. Se dice también que en la memoria a corto plazo se retiene ante todo las propiedades físicas de la información, mientras que en la memoria a largo plazo lo es especialmente la información semántica, aunque esta última clase de memoria se caracterice por la enorme variedad de conocimientos que es capaz de albergar.

Entre las casillas de las memorias a corto y a largo plazo se dibujarán flechas en las dos direcciones para dar a entender las interacciones que tienen lugar entre ambas clases de estructuras cognitivas. Pues, en efecto, se ha propuesto que, en determinadas ocasiones, tiene que haber un transvase de la información sensorial retenida brevemente en la memoria a corto plazo a la memoria a largo plazo, y viceversa. Por lo demás, si este modelo, que se apoya en la concepción de la memoria a corto y a largo plazo como dos estructuras rígidas diferentes, no es del todo adecuado, como tienden a mostrar algunas investigaciones recientes, puede postularse la existencia de un único sistema de memoria con distintos estados funcionales. Pero, como digo, los detalles, aunque importantes, no son aquí de demasiado peso.

Aparte de estas dos estructuras cognitivas básicas, se destaca el hecho de que en los procesos de análisis perceptual y en otras tareas cognitivas complejas, la atención tiene un papel predominante. La atención, considerada desde el punto de vista de los análisis de la psicología cognitiva, es un *mecanismo* de carácter selectivo y de capacidad limitada que funciona en muchas ocasiones, a su vez, como mecanismo endógeno de alerta ante ciertos estímulos ambientales (Vega 1984, 124). La atención es, por otra parte, un proceso activo que permite seleccionar una parte de toda la información que concurre en el sujeto en un determinado momento para que sea sólo esa parte, o predominantemente esa parte, la que reciba un mayor procesamiento por parte de los recursos cognitivos de que dispone el organismo. Desde luego, no ha escapado a la reflexión de los psicólogos

cognitivos que la atención, por sus propiedades peculiares, es uno de los fenómeno psicológico que más cerca está de las manifestaciones de la conciencia. Así, Manuel de Vega, cuyo manual de psicología cognitiva me está sirviendo para elaborar este mapa general de los procesos cognitivos, comenta a propósito de las relaciones entre atención y conciencia:

Uno de los términos más desacreditados de la psicología científica tradicional es el de conciencia. Muchos psicólogos actuales de orientación científica rechazan este término por considerarlo cargado de connotaciones fenomenológicas y mentalistas. Sin embargo, la psicología cognitiva ha terminado por recuperar la idea de conciencia, que se considera como correlato subjetivo de los procesos atencionales... *La mayoría de los modelos teóricos que presentaremos... son elaboraciones mecanicistas basadas en datos conductuales y no en informes introspectivos*; sin embargo, las alusiones a la experiencia consciente de los sujetos resultan a veces inevitables. Por ejemplo, cuando el sujeto recibe simultáneamente dos mensajes verbales y atiende selectivamente a uno de ellos, *se manifiesta en ciertos índices objetivos de rendimiento*, pero también resulta natural hablar de que el sujeto es *consciente* de uno de los mensajes (el seleccionado), ya que su experiencia subjetiva de éste es evidentemente muy distinta, de la del mensaje rechazado.

... La atención sería el fenómeno subyacente a una serie de rasgos conductuales cuantificables y al fenómeno experiencial de la conciencia. (Vega 1984, 125-126. Los subrayados son míos).

Si hay textos importantes en los que bucear para sacar a la superficie los presupuestos teóricos de la psicología cognitiva, así como para rastrear la ontología de la conciencia que, todavía más oculta que aquellos presupuestos, no puede, sin embargo, dejar de presentirse a cada momento, éste es, desde luego, uno de ellos. Para el psicólogo, la atención del sujeto es, antes que nada, algo que "se manifiesta en ciertos índices objetivos de rendimiento" respecto de las tareas experimentales propuestas. El texto indica, por otra parte, la *naturalidad*--¿propensión?--con que hablamos, cuando tenemos delante de nuestra estimación teórica esos mismos datos,

de que el sujeto es consciente de esto o de lo otro con diversos grados de intensidad atencional. Por lo demás, estas líneas están plagadas de algunas expresiones misteriosísimas para quien se acerque por primera vez a estos problemas. Pues, ¿acaso no es necesario decir que no se comprende completamente, entre otras cosas, qué significa eso de que "la atención sería el fenómeno subyacente a una serie de rasgos conductuales cuantificables"? Sin embargo, desenmarañar el camino de pensamiento que puede ayudar a comprender mejor qué se quiere decir con estas palabras no es nada fácil. Quizás, lo mejor para una mente entrenada en el análisis es proceder por partes. Propongo, en este sentido, ofrecer primero una explicación de cuáles son las reticencias que los psicólogos han encontrado para ocuparse de la conciencia y de la atención desde el punto de vista fenomenológico con el fin de ver que hay ahí que es tan digno de descrédito y por qué lo es. Esta labor preliminar es, además, de una importancia extraordinaria para la discusión de las teorías de Dennett.

¿Por qué se considera que el fenómeno de la atención está tan íntimamente ligado a la idea de conciencia? Para entender este punto debidamente lo mejor que se puede hacer es echar una ojeada a los análisis fenomenológicos y, precisamente, a aquellos que han despertado y despiertan tantas susceptibilidades entre los psicólogos de "orientación científica". El motivo para proceder así—meterse directamente en la boca del lobo—es que la fenomenología husserliana fue la primera que elaboró una teoría descriptiva algo detallada en la que atención y conciencia no podían desligarse en absoluto. Si la psicología cognitiva ha recuperado en algún sentido la noción de conciencia, pero ha logrado hacerlo desprendiéndose ampliamente de las connotaciones fenomenológicas que tenía, conviene saber entonces cuáles son estas connotaciones fenomenológicas y de qué manera se podría alguien alejar de ellas.

Los fenómenos de la atención tienen la peculiaridad, desde el punto de vista fenomenológico, de acompañar en todo momento nuestra vida de vigilia. Estar despiertos casi parece ser sinónimo de atender ahora a esta tarea, ahora a esta otra, ahora a esta música, ahora a esta escena en la calle, etc. Claramente, el primer lugar donde se puede encontrar este fenómeno peculiar es en cada uno de nosotros. Muchos estudiantes preparan exámenes escuchando música. Sin embargo, si están concentrados en sus estudios, es decir, si "vuelcan" toda su atención sobre el contenido de sus apuntes, y tratan de reproducir un razonamiento o memorizar un determinado texto, la música les parecerá que ha desaparecido, como si no existiese, cuando se les pregunta posteriormente sobre ello. Después de permanecer un tiempo concentrados, la atención sobre el libro o los apuntes parece que de pronto se distiende. La aparición del más ligero desánimo--esto "no me entra", esto es muy difícil, ya estoy cansado--ocasionará que uno "se despegue" del libro. En ese momento, el estudiante seguramente se revolverá sobre la silla, dejará el bolígrafo sobre la mesa, se rascará la cabeza y estirará el cuerpo. Quizás justo también en ese momento le parecerá que la música suena de nuevo por ahí detrás, no del todo atendida, pero tampoco del todo desatendida, mientras que también sigue "zumbando" simultáneamente en su conciencia el contenido del texto que quería memorizar. Basta, por lo demás, con que la radio esté emitiendo una de sus canciones favoritas para que de inmediato se sienta irremediablemente "atraído" por ella, se desentienda de todo lo demás que estaba haciendo y empiece a atender plenamente a la música y a tararearla.

Este ir y venir de la atención de una actividad a otra, de memorizar un texto a cantar una canción, y de hacer cada una de estas actividades con diversos grados de atención, son casos de lo que llama Husserl genéricamente "desplazamientos atencionales" (*intentionale Wandlungen*) (1950, 228). Por otra parte, para aclarar también la terminología que aparecerá

posteriormente, el texto atendido para su memorización y la canción de pronto atendida son ejemplificaciones posibles del concepto husserliano de "nóema". Son contenidos noemáticos. Por su parte, la actividad de memorizar el libro o de canturrear la canción son, en la terminología técnica de la fenomenología, distintas especies de "nóesis" igualmente posibles.

Husserl, ciertamente, se expresa en muchas ocasiones de una manera completamente metafórica, aunque consciente de estar haciéndolo así, cuando habla de los diversos fenómenos atencionales. Se sirve de expresiones tales como "la mirada del espíritu" o el "rayo visual" del yo puro que atraviesa ahora esta nóesis con este contenido noemático concreto y ahora esta otra (1950, 228). Por ejemplo, la mirada del yo que atraviesa ahora esta vivencia continua que consiste en la actividad de memorizar este texto, o la mirada del yo que atraviesa simultánea, o alternativamente en otro ahora, el recuerdo y la ejecución de esta canción. Esta es una de las especies posibles del desplazamiento atencional. La otra especie de desplazamiento atencional no es la que se mueve completamente de una especie de nóesis a otra y, correspondientemente quizás, de un nóema a otro. Antes bien, consiste en desplazamientos de la intensidad con que el yo atiende hora a uno, hora a otro nóema. La música pasó de "estar como muerta para la conciencia", a "estar coatendida" junto con el recuerdo de lo que estaba estudiando, para, finalmente, "estar atendida primaria y exclusivamente" ella sola. Todas las expresiones entre comillas son expresiones literales o paráfrasis casi literales de nociones que utiliza el propio Husserl (1950, 229-230). En general, Husserl divide todos estos grados en las intensidades atencionales, que constituyen una de las especies posibles de desplazamientos atencionales, en dos extremos a los que llama respectivamente, modos de la actualidad (Aktualitätsmodi) y modos de la inactualidad (Inaktualität) (1950, 230).

Las descripciones de Husserl del fenómeno de la atención, en tanto que "estructura de todo punto universal de la conciencia" (1950, 228), le llevan a considerar que el campo de la conciencia está como segmentado en capas más o menos concéntricas--como las de una cebolla--respecto de un núcleo central verdaderamente atendido al que denomina "núcleo atencional" (das attentionale Kern) (1950, 230). Este núcleo atencional sería el propiamente habitado de una forma completa por la mirada consciente del yo. Para Husserl, toda vivencia así atendida es a lo que habría que llamar realmente una cogitación del yo. Alrededor de este centro se irían desplegando sucesivamente diversas capas más o menos coatendidas hasta llegar al último margen en el que se daría la pura afección, el momento en el que algo "entra" en el campo de la conciencia, pero en el que ésta se mantiene completamente pasiva respecto de aquello que ha aparecido. Así, pues, y de acuerdo con esta nueva descripción, es posible dividir todas las referencias intencionales en dos modos según sea su relación al yo. Referencias intencionales en el modo de la actualidad del yo (Ichaktualität), y referencias intencionales alojadas en el transfondo consciente del yo (Bewusstseinshintergrund des Ich). A este trasfondo de la conciencia lo denomina Husserl también corte (Hof) de la conciencia y constituye el campo de su libertad (Feld seiner Freiheit). Respecto de los modos de la actualidad, limitados por el grado superior de la atención, dice Husserl que están investidos por "*el carácter de la subjetividad*, (de la yoidad) (Ichlichkeit)" (Husserl 1950, 231), y es propio de esta subjetividad el poder desplazarse con más o menos libertad por el campo de los nóemas más o menos atendidos y coatendidos que están en su radio de acción. A su vez, esto no significa nada más que la posibilidad que tiene el yo de atender ahora a esto, ahora a aquello--al texto o a la música--, e incluso de padecer simplemente la aparición de uno u otro contenido--de no poder substraerse,

por ejemplo, a la atracción de su canción favorita o a la aparición de un ruido estruendoso.

Aquí terminaría una primera descripción y caracterización general de las relaciones entre atención y conciencia desde el punto de vista fenomenológico. Ahora, según el psicólogo, es momento de preguntarse: muy bien, ¿y qué? Todo esto es muy bonito, entretenido y ciertamente interesantísimo y esclarecedor. Requiere seguramente, además, una gran sutileza y esfuerzo serio y constante poder llegar a hacer una descripción de este estilo y encontrar los términos precisos para plasmarla adecuadamente en palabras. Incluso, es en descripciones de ese estilo donde se muestran, si se quiere "originariamente", los fenómenos psicológicos que el psicólogo quiere estudiar empíricamente. Pero el psicólogo, a partir de aquí, quiere saber todavía muchas más cosas, pues, desde su punto de vista, una descripción de este estilo no hace avanzar ni un ápice nuestro conocimiento de psiquismo humano. Las preguntas a partir de aquí son, más bien, estas otras: ¿cómo y por qué así?, ¿qué mecanismos son los que subyacen a estos desplazamientos atencionales que constriñen la capacidad atencional del yo?, ¿cómo operan?, ¿cómo los fenómenos atencionales se pueden poner en relación con otros fenómenos psicológicos y con otros datos disponibles para el psicólogo? Por su parte, el filósofo de orientación metateórica estaría interesado en saber, por ejemplo, en qué sentido habría que distinguir el fenómeno de la atención respecto de estos mecanismos que postula la teoría psicológica.

--¡Ah! esto ya no es asunto mío. Esto es una cuestión de la psicología empírica, aunque sea, desde luego, una cuestión legítima-- responde el fenomenólogo--. Yo lo único que digo es cómo me parece indudablemente que hay que realizar estas descripciones cuando me atengo exclusivamente a la forma en que se me dan las distintas especies de desplazamientos atencionales desde el punto de vista fenomenológico, desde

el punto de vista de la "descripción inmanente" de esos fenómenos tal y como se me dan en la conciencia. Además sostengo que sólo cuando realizo esta descripción es cuando se me da originariamente y entiendo el concepto general de atención del cual, por otra parte, es imposible hablar sin hablar del yo que atiende a esto o a lo otro. El resto es naturalizar lo psíquico y, en verdad, en este sentido exacto:

Al alma le es asignado un tipo de ser principalmente análogo al de la naturaleza, a la vez que a la psicología se le asigna la tarea de elevarse teóricamente desde la descripción a la última "explicación" teórica, como en la biofísica. (Husserl 1976a, 64).

Probablemente el fenomenólogo tenga, en parte, razón. En cualquier caso, tenemos aquí de nuevo un caso más de la división del trabajo antes señalada. El filósofo ya ha cumplido con su tarea y ahora es el turno del psicólogo interesado en hacer teorías empíricas de los distintos fenómenos psicológicos. Supongo que Dennett se dirá: ¡ay estos chicos, siempre igual, cada uno por su lado!

¿Y qué dice el psicólogo? Muchas cosas y muy interesantes. Es importante comenzar por algunos hechos "externos". Lo primero que puede observarse cuando se tiene entre las manos un libro de psicología cognitiva es que después de los primeros pasos por las descripciones y los ejemplos más comunes del fenómeno psicológico que se quiere investigar se da un giro radical en la terminología que se emplea. (La justificación de esta terminología se suele ofrecer en los primeros capítulos.) Al hablar, por ejemplo, de la atención, se utiliza al principio expresiones metafóricas intuitivas como "el foco de la atención" (Vega 1984, 124). Todas estas expresiones se reemplazan posteriormente por otras expresiones que tienen su origen en el vocabulario técnico del procesamiento de la información.

La atención es, por tanto, un mecanismo de control activo que permite al procesador una toma de posición ante los *inputs*. (Vega 1984, 124).

Sin embargo, no parece que decir una cosa de este estilo suponga un gran avance respecto de las descripciones intuitivas o metafóricas. Si se sustituye "procesador" por "sujeto" e "*inputs*" por "nóemas" parece que estamos en las mismas. En el caso de Dennett se puede encontrar también una dificultad semejante. La única diferencia es que él llama "control" a lo que Vega llama "procesador" (Dennett 1981, 156-157). De cualquier forma, da la sensación de que en el modelo de procesamiento de la información tanto el "procesador" como el "control" adquieren las funciones que intuitivamente atribuimos al yo, a nuestro yo. Cuando se presentan así las cosas da ganas de preguntarse inmediatamente dónde están entonces las ventajas del modelo de procesamiento de la información para el tratamiento de los fenómenos psicológicos. Quizás una mente ágil y rápida vaya incluso más allá y proponga que si no hay respuesta adecuada a esta pregunta, entonces no es posible un tratamiento subpersonal ni de este fenómeno psicológico esencial, ni de ningún otro. Como se verá más adelante, la solución final que propone Dennett para este y para otros fenómenos psicológicos conscientes es dar la vuelta a la tortilla al planteamiento global del problema. No obstante, antes de recorrer las distintas etapas de esta posible y atrevida solución conviene decir algo más acerca del problema concreto de la atención.

Lo primero que podría hacer un psicólogo o un científico cognitivo para legitimar el uso de la terminología técnica del procesamiento de la información es, sin duda, apelar al hecho de que los modelos computacionales proveen marcos explicativos más potentes en relación a las meras descripciones fenomenológicas. Estos marcos son más adecuados porque ofrecen la posibilidad de integrar de una manera fructífera y

racional numerosos datos relacionados con el fenómeno psicológico que se quiere estudiar. La investigación puramente fenomenológica no parece que ofrezca semejantes ventajas.

Ya las primeras investigaciones empíricas de Cherry y Broadbent del fenómeno de la atención y sus primeros modelos teóricos dieron lugar a hablar de la atención como de un mecanismo que actuaría de una forma análoga a un filtro rígido. Estas primeras investigaciones de los años cincuenta, fundadas sobre todo en experimentos con mensajes acústicos dicóticos, confirmaron algunas propiedades intuitivas de la atención como, por ejemplo, que no se puede realizar con eficacia dos o más tareas que supongan una gran dificultad para el sujeto. Pusieron también de relieve otras propiedades no tan accesibles para nuestro modo de pensar cotidiano. Así, que de dos mensajes, uno se puede procesar completamente y el otro no, aunque este último reciba también procesamiento en sus propiedades sensoriales más básicas (Vega 1984, 128-131).

Después de estas investigaciones pioneras, ha habido muchas discusiones entre los psicólogos para determinar si el modelo de filtro rígido de la atención es o no adecuado. También se ha desatado la polémica para establecer si el filtro, en caso de existir, actuaría de una forma precategorial sobre los estímulos sensoriales almacenados o si, por el contrario, actuaría de una forma poscategorial. Respecto de esta última cuestión, el modelo de filtro rígido de Broadbent proponía que toda la información sensorial que concurre en el sujeto en un tiempo dado se procesa en paralelo, esto es, simultáneamente, y se retiene, de una forma relativamente muy poco interpretada, en la memoria a corto plazo a la que Broadbent llamaba "memoria sensorial". Para poder analizar el mensaje en sus rasgos más complejos-sintácticos o semánticos, por ejemplo-supuso Broadbent que tenía que haber un filtro selectivo que, de entre toda la información sensorial que el organismo recibe simultáneamente por los

sentidos, dé entrada a cierta sección de toda esta información. Es esta parte la que, una vez atravesado el filtro, lograría ser procesada y analizada más detalladamente en sus rasgos más complejos y la que "el sujeto", finalmente, expresaría como un contenido consciente suyo. El comentario necesario para tomar las debidas precauciones no hace falta casi ni mencionarlo. Que el filtro actúe para seleccionar uno u otro mensaje depende, obviamente, de determinadas propiedades de los estímulos--intensidad, localización, sentido por el que se recibe, velocidad de presentación--así como del estado y de las necesidades del propio organismo. Se reconoce, por consiguiente, que la atención, en tanto que mecanismo psicológico, no sólo viene determinada por las propiedades de los estímulos, sino también por las expectativas del propio organismo. En resumidas cuentas, que la atención es tanto un proceso guiado de abajo-arriba (data driven processing) como guiado de arriba-abajo (conceptually driven processing).

Como he anunciado, contra este modelo precategorial de la atención se ha alzado un rival que parece explicar más datos. La pregunta que ha suscitado que los investigadores se pongan a buscar un modelo alternativo al del filtro rígido ha sido ésta: "¿cómo es posible seleccionar un *input* de entre otros concurrentes, a menos que se analicen todos ellos?" (Vega 1984, 135). Este argumentos lógico, unido a multitud de otros datos experimentales, ha dado lugar a la aparición del modelo del filtro poscategorial. La diferencia básica de este modelo respecto del anterior es que las señales procedentes de los órganos de los sentidos se procesan en paralelo--mucho más allá, por otra parte, del procesamiento de sus propiedades físicas--*antes* de que entre en acción el filtro. El filtro tiene ahora la función de recoger y evaluar las señales analizadas previamente así como de seleccionar el mensaje relevante. Es este mensaje que se ha seleccionado el que se alberga finalmente en el "*mecanismo central de*

atención" y el que constituye el contenido del estado consciente del sujeto (Vega 1984, 136).

Este último modelo de la atención explicaría mejor que otros los datos empíricos obtenidos en el experimento de Lackner y Garrett que se analizó en el capítulo segundo con el que se pretendía construir un modelo que diera cuenta de cómo es posible que las oraciones no atendidas influyan decisivamente a la hora de parafrasear el sentido de la oración atendida. Dennett, haciendo uso también del análisis de este experimento psicológico clásico, y aplicando los resultados en él obtenidos a las distinciones antes mencionadas entre acceso computacional, público y consciente a un contenido, propone las siguientes consideraciones. El modelo de filtro rígido precategorial de Broadbent no funciona bien como hipótesis para dar cuenta de los datos de este experimento. Como se ha expuesto, este modelo sólo asigna un nivel de procesamiento superior-sintáctico y semántico-a aquellos estímulos que han pasado ya el filtro atencional. El resto de las señales quedarían sin procesar en sus propiedades semánticas y sólo lo serían, en todo caso, en sus propiedades físicas más básicas (voz de mujer o de hombre, mero ruido o música, etc.). En este sentido, las propiedades semánticas de las señales no atendidas no se analizarían en absoluto. Los datos empíricos recogidos escrupulosamente por Lackner y Garret sugieren, por el contrario, que también el mensaje no atendido debe de haber sido procesado hasta sus propiedades semánticas, aunque el sujeto no sea consciente de ello.

Haciendo pie en este experimento, Dennett extrae la misma conclusión que extraía Fodor acerca de la poca evidencia que merece en algunos casos la reflexión del sujeto sobre sus propios contenidos conscientes, sobre todo si es que esta reflexión quiere tomarse como la única evidencia que debe ser tenida en cuenta por la psicología.

La *ausencia* de evidencia introspectiva de que se ha realizado cierto análisis no es nunca evidencia fiable de que semejante análisis no se ha realizado (Dennett 1981, 159).

Es ahora cuando se puede poner en juego por primera vez las nociones de acceso computacional, acceso público y acceso consciente a un contenido para verlas en su sentido. El acceso computacional que tienen los distintos sistemas de procesamiento a los mensajes que ellos analizan, es independiente del acceso público a esta información y del acceso consciente del sujeto. En el nivel computacional es donde se llevaría a cabo tanto el análisis semántico de los dos mensajes concurrentes como la "comparación" semántica del contenido del mensaje atendido con el del desatendido. Si no malinterpreto todo el asunto, lo que Dennett sugiere en este caso es que la atención consciente del sujeto a uno de los canales no influye para nada en todos estos procesos. Su acción simplemente consiste en mantener activada la subrutina de impresión que accedería computacionalmente al *resultado* del procesamiento preconscious del contenido semántico de los dos mensajes. Es ese resultado, que se supone que queda albergado en algún lugar del sistema cognitivo, el que se "publicaría" posteriormente gracias al sistema lingüístico del sujeto. Y es solamente a este resultado final al que tendría acceso directo el sujeto de una forma consciente. A él también accedería indirectamente el psicólogo gracias a lo que el sujeto dice y hace, gracias, por consiguiente, a ciertos "índices objetivos de rendimiento" que tienen un significado preciso en la situación experimental heterofenomenológica.

Teniendo en cuenta estos elementos teóricos y metateóricos, la relevancia de toda esta discusión para el problema que se planteaba a propósito de la justificación de los modelos de procesamiento de la información así como del vocabulario técnico que llevan aparejados, se deja advertir cuando se considera que la sistematización racional de los datos

parece reclamar tanto estos modelos como su peculiar vocabulario. Ahora, bien, estos modelos sugieren abiertamente la posibilidad de hablar de las operaciones de la atención como de una mera "distribución de recursos cognitivos" (allocation of cognitive resources) (Dennett 1981, 157). Dado que los recursos cognitivos del sujeto son limitados, la atención funcionaría como el mecanismo que repartiría estos recursos según las tareas que se propongan al organismo. Dennett hace las siguientes declaraciones cuando comenta el mecanismo de la atención así entendido y su posible relación con la conciencia:

Esta noción de atención, pese a su importancia, está sólo muy indirectamente conectada con la conciencia, como puede verse inmediatamente si uno considera el hecho de que cualquier computadora que soluciona problemas o que juega pone atención, en este sentido, primero a un curso de acción elegible y luego a otro, y es de suponerse que no por esta razón sería considerada consciente... En este sentido de atención, *atención inconsciente* no es una contradicción en los términos y, de hecho, no se ha dado ninguna indicación que sugiera lo que podría ser la atención *consciente* (1981, 157).

Como se ha puesto de relieve en este párrafo, sí que se han dado indicaciones precisas--o, al menos, ciertas descripciones más o menos elaboradas--acerca de qué sería la atención consciente. Lo que ocurre es, supongo, que no son indicaciones que sean tratables por la psicología cognitiva a no ser que el psicólogo las incluya de alguna manera en su estrategia general en relación a todas las manifestaciones posibles de un yo, de una conciencia, a saber, como aquello que precisamente hay que explicar en sus propiedades y estructuras y que se manifiesta en datos conductuales objetivos que son recabados de una forma óptima en condiciones de laboratorio.

La importancia enorme que tienen todas estas consideraciones para la teoría de la conciencia y del sujeto puede ser fácilmente adivinada. De todos estos problemas se hablará en breve. De momento, para abrir bocado, Dennett dice lo siguiente:

... nosotros simplemente *podemos decir* qué es lo que estamos experimentando. Esto se consigue sin ningún ojo interior o facultad introspectiva aparte de la maquinaria invocada en el modelo (Dennett 1981, 170-171).

Es decir, no hace falta ninguna facultad introspectiva, ni rayo del yo que atraviesa nada, ni ojo interior, ni cosa que se le parezca. Simplemente existe esa maquinaria natural que sería lo único realmente excelente aquí. Por muy maravilloso que nos parezca que un sujeto tenga conciencia de algo a lo que presta atención, ni siquiera eso es lo suficientemente maravilloso como para que no pueda ser producido por el mecanismo de la naturaleza, como decía la cita de Leibniz que preside este capítulo. Al menos, esto es lo que piensa Dennett.

& 4. Heterofenomenología de la conciencia.

Después de haber analizado con cierto apresuramiento—aunque espero que no de una forma completamente insuficiente—algunos de los distintos subsistemas que formarían parte del organigrama funcional general que representaría idealmente al psiquismo humano, la pregunta que pende como una plomada para corregir y mostrar las posibles deficiencias de las paredes de todo el edificio psicológico cognitivo así construido es: ¿acaso se puede decir de él en algún sentido que es consciente? Para muchos el edificio ha perdido toda su verticalidad y amenaza peligrosamente con derrumbarse porque la conciencia no se ha empleado prácticamente para nada en su

construcción. ¡Pero si no hay más que verlo, si está todo él que parece que se cae! Para otros, por el contrario, su posible inclinación forma tan sólo parte de un proyecto muy bien trazado y calculado, y no hay ningún peligro de derrumbamiento, aunque sus paredes no estén rectas respecto de determinados varemos de la arquitectura más clásica. Lo que importa es que el producto final no se caiga. El edificio, a decir verdad, es sólo una muestra representativa, sólida y audaz de la arquitectura contemporánea. En este terreno las plomadas para calcular la verticalidad están obsoletas. Otros cimientos y otros materiales dan como resultado otros edificios y otros varemos.

Dado que las ciencias cognitivas parecen apañárselas muy bien con las nociones de acceso computacional y público a la información, la idea de un yo consciente, del rayo del yo que vive sus vivencias en el modo de la actualidad y las convierte propiamente en sus *cogitationes*, no disfruta de buena salud a no ser como descripción refinada de nociones de sentido común que están plagadas de metáforas bonitas, pero finalmente poco útiles para la labor explicativa teórica. Para la psicología científica, en tanto que ciencia egológica empírica, estas descripciones y estas metáforas están, hasta cierto punto y respecto de sus fines, fuera de lugar a no ser como punto de partida inteligible. Pero, ¿qué otros fines se podrían tener? ¿Qué utilidad tendrían estos otros fines para el conocimiento de la mente y del psiquismo humano?

Atención porque ahora viene lo bueno. Que todo el mundo se ponga el casco por si acaso. Después de todo este recorrido por la obra en construcción, debería resultar claro que el psicólogo empieza enfrentándose con un ser supuestamente consciente, con un ser psicológicamente semejante a él, pero que termina, cuando quiere explicar la conducta de este sujeto, con un sistema funcional complejísimo formado de partes subpersonales completamente oscuras para el ojo de la conciencia

del sujeto de experimentación. El filósofo de orientación metateórica, por su parte, asegura y apoya estos mismos resultados. De hecho, no podía hacer otra cosa.

Con estos elementos a la vista, el giro que da Dennett a toda el problema de la conciencia es doble. La primera vuelta consiste en adoptar el punto de vista heterofenomenológico. Para el psicólogo que mira al sujeto de experimentación "desde fuera", desde la exterioridad de su conducta, este sujeto pasa de ser un ser consciente a ser aquella entidad funcional sumamente compleja a la que se hacía referencia en el modelo que se ha presentado. La situación global es así descrita por el propio Dennett:

...nosotros en el exterior, al mirar tal entidad, al conversar con ella, al escuchar sus esfuerzos por describir los efectos en ella de varios ambientes perceptuales, tendremos por lo menos la ilusión de que se siente algo ser la entidad. De hecho, ella nos dirá (o al menos parecerá que nos dice) exactamente qué se siente. Pero en el interior todo es oscuridad, engaño. (1981, 164-165).

La segunda adopta más directamente aún el tono de la provocación. Admitamos la hipótesis de trabajo de inspiración "platónica", completamente normal en las ciencias especiales, de que el psiquismo humano, por lo que respecta exclusivamente a los procesos cognitivos, presentará en la asíntota de la investigación el aspecto de un organigrama muy complejo, y que la tarea de las ciencias cognitivas es dar con dicho organigrama y explicarlo en sus propiedades, estructuras y relaciones esenciales. Supóngase también, correspondientemente, que "propongo la hipótesis audaz de que usted es una realización de este diagrama de flujo, y de que es en virtud de este hecho que parece--que nos parece a nosotros y a usted--que se siente algo ser usted. ¿Podría dar buenas razones para negar la hipótesis y, si es así,

cuáles son? (Dennett 1981, 165). Algunas páginas después vuelve a comentar:

Cuando nos preguntamos si algo o alguien es consciente, es tentador ver esto como si preguntáramos si hay o no una luz especial encendida dentro. Esto es, sin embargo, un error como podemos ver al hacernos preguntas sobre nuestro propio caso: ¿era yo consciente (o consciente de X) en el tiempo t ? Cuando vemos que lo que afirma las conclusiones en nuestro propio caso es una consideración de hechos relativos a nuestras capacidades presentes y nuestras actividades pasadas y la mejor teoría que puede dar razón de ellas, nos mostramos menos renuentes a admitir que las mismas consideraciones decidan el asunto en el caso de los demás. (Dennett 1981, 172).

& 5. La conciencia y el texto psicológico de Dennett.

Es importante volver sobre los pasos ya andados. El psicólogo tiene una forma muy peculiar de acercarse al estudio de la conciencia del sujeto de experimentación y a los distintos fenómenos psicológicos en los que ella se manifiesta. Para ser lo suficientemente precavido y no comprometerse demasiado con una caracterización determinada de la conciencia antes de emprender toda investigación psicológica empírica posible, el teórico la podría considerar simplemente como una entidad de la que proceden o pueden proceder ciertas respuestas que son valoradas *eo ipso* como datos experimentales fiables. Pero, ¿acerca de qué versan estos datos? El psicólogo se dice a sí mismo más o menos calladamente: quiero recoger de una forma fiable--en condiciones máximamente fiables--el mayor número posible de datos empíricos que sean pertinentes para hacer una teoría empírica de la percepción, de la atención, o de la memoria, o de la resolución de problemas, o de la interacción entre dos o más partes del posible sistema que vaya emergiendo de estos estudios parciales. Por ejemplo entre la atención y la percepción. Para este fin, mucho más importante aún para los fines de la fundamentación de la psicología como ciencia egológica

empírica que señalar que toda acumulación e interpretación de datos empíricos está mediada esencialmente por algún tipo de teoría--que no hay observación pura, como suele decirse--, me parece que es constatar claramente que el psicólogo es el *origen último* que establece la relación entre los datos observacionales y el fenómeno psicológico que quiere estudiar. Alguna noción más o menos clara de lo que es una percepción o un fenómeno atencional debe de tener previamente el psicólogo cuando se dice que va a hacer una teoría de la percepción o de la atención basada en datos empíricos. El supuesto de que el sujeto de experimentación es psicológicamente semejante a él juega también un papel importante, como hubo ocasión de ver, para construir el experimento psicológico que rendirá los datos empíricos deseados que confirmen la teoría. Si yo percibo o atiendo y entiendo qué es percibir o atender, entonces mi sujeto de experimentación también percibirá o atenderá de una forma muy semejante a mí, y si yo le pongo hábilmente en una situación experimental adecuada que me permita el control de determinadas variables relevantes para el estudio de la percepción o de la atención, entonces obtendré determinados datos que, en el mejor de los casos, confirmarán mi hipótesis previa, la hipótesis que ha guiado precisamente la recolección e interpretación de esos datos, así como su conexión sistemática en el todo de la teoría. Habré obtenido una teoría empírica subpersonal de la percepción o de la atención. Ahora bien, si yo, como psicólogo, percibo y puedo decir, por consiguiente, "yo percibo conscientemente esto o lo otro", entonces puedo suponer que mi réplica psicológica también podrá decir "yo percibo conscientemente esto o lo otro". Sin embargo, mi investigación empírica de la percepción se quedará solamente con su "esto o lo otro" en forma de dato empírico psicológicamente relevante desdeñando, en parte, su "yo".

Como se ha visto en varias ocasiones, el "esto o lo otro" del sujeto de experimentación será expresado típicamente por dicho sujeto mediante

un informe verbal más o menos preciso. Pero no solamente utilizando informes verbales. Podrá expresarlo también de muchas otras formas. Pulsar un determinado botón de acuerdo con instrucciones previas fijas y bien establecidas--el de la derecha significa "sí" el de la izquierda "no"--valdrá igualmente para expresar su "esto o lo otro" como actos de habla estereotipados en una convención. Por lo tanto, en el experimento psicológico lo más relevante no es el cómo se representa el "esto o lo otro", sino que lo esencial es el qué se representa (Dennett 1982b, 165-166).

Aparte de la expresión y representación, verbal o no, del contenido consciente--del "esto o lo otro"--habrá en numerosos casos *otro* conjunto de datos empíricos aportados por la situación experimental en los que no se recoge ya la información que se pide al sujeto que aporte explícitamente cuando se le asigna una determinada tarea. Antes bien, son datos experimentales "adyacentes", pero que también se consideran de excepcional importancia para construir paso a paso y de una manera sólida la teoría. Serán, por ejemplo, el registro de las frecuencias de onda que emite el cerebro del sujeto cuando está realizando la tarea experimental propuesta, la medida de los tiempos de reacción entre la aparición de los estímulos relevantes y la aparición de la respuesta consciente del sujeto, las oscilaciones galvánicas registradas en la piel del organismo y un largo etcétera. Estos dos grupos de datos, los que aporta explícitamente el sujeto de experimentación y los "adyacentes", son los que quiere recoger escrupulosamente el psicólogo cognitivo mediante un trabajo largo, duro e intensivo. El psicólogo, y esto es esencial, pone estos datos en conexión más o menos inmediata con los estados perceptuales, o atencionales del sujeto, o con cualesquiera que sean los fenómenos psicológicos que esté estudiando. La relevancia que tiene proceder de esta manera es que "cualquier cosa que sea lo que estemos estudiando, y lo bien o mal que lo estemos estudiando, lo estaremos estudiando empíricamente" (Dennett

1982b, 160). La acumulación de datos empíricos, en una acepción muy amplia de esta expresión, siempre servirá para la investigación empírica de algo que no son esos mismos datos empíricos.

Del conjunto de todos los datos recogidos por el psicólogo dice Dennett que forman un *texto*. El texto que el psicólogo tiene delante de sus manos son los datos recabados por él cuando se propone como tarea la construcción de una teoría sobre un determinado fenómeno psicológico que esté basada en datos empíricos. Es el psicólogo, desde luego, el que ha de interpretar dicho texto de acuerdo con ciertos presupuestos heurísticos².

El instrumento heurístico básico con el que cuenta el psicólogo para interpretar su texto lo constituye la adopción, desde el inicio mismo, de la *instancia intencional*. Adscribir estados mentales a un organismo es fundamental cuando queremos situarnos inicialmente en el nivel de descripción adecuado para explicar racionalmente su conducta. Enfrentado con su texto, el psicólogo considerará sencillamente que su existencia "es el producto de un proceso que tiene una *interpretación intencional*: consiste en cosas que el sujeto quería decir, en proposiciones que quería afirmar por varias *razones*" (Dennett 1982b, 161). Creo que esta interpretación intencional no es nada más que otra forma posible de decir que el psicólogo considera en todo momento al sujeto de experimentación como un ser psicológicamente semejante a él. Con toda seguridad, el psicólogo se "interpreta" a sí mismo intencionalmente y extiende esta interpretación

² Una detallada y aguda comparación entre los métodos y resultados de la hermenéutica en el estudio e interpretación de los textos literarios, con los métodos y resultados de la psicología empírica cuando se enfrenta a sus propios "textos", le sirve muy bien a Dennett para hacer fuerte o, al menos, muy plausible, la posición que defiende que, en el fondo, se está haciendo lo mismo y de la misma manera en las ciencias humanas que en las ciencias naturales, aunque con diferentes objetos. Esta comparación, por su parte, valdría para establecer que en determinados momentos las ciencias humanas pueden tender una mano amiga a las ciencias naturales.

a todo ser que considere psicológicamente semejante a él en algún rasgo relevante.

Cuando se ha dado este primer paso en la interpretación del texto psicológico se plantea inmediatamente la consideración de si esta interpretación intencional no viene determinada *necesariamente* por el hecho de que el autor del texto es un ser consciente como lo es el psicólogo. Al fin y al cabo, este último parece partir del presupuesto de que existe una conexión entre las diversas especies de estados psicológicos que quiere investigar y los datos que el sujeto de esos estados psicológicos aporta en la situación experimental. Creo que la respuesta de Dennett a esta pregunta constituye la pieza clave de toda su teoría de la conciencia: *no existe ninguna conexión necesaria entre la conciencia y la interpretación intencional*. Por consiguiente, si se establece alguna conexión en este sentido, dependerá completamente de una elección no forzada por parte del psicólogo (1982b, 163). Es decir, que la tesis de la semejanza psicológica no obliga a hablar del sujeto de experimentación como de un yo que percibe esto o aquello, que recuerda esto o aquello, sobre todo si es que se tiene la propensión a considerar este yo como algo determinable y analizable al margen de la investigación psicológica teórica y empírica concreta.

¿Con qué razones apoya Dennett esta afirmación tan sorprendente? Para defender este punto de vista Dennett llama en su ayuda a la fina comparación que mencionaba más arriba y que se establece entre los métodos de la hermenéutica para la interpretación de los textos de ficción literaria y los métodos que podría adoptar el psicólogo en relación a su texto psicológico. En una novela, por ejemplo, la narración habla acerca de determinados personajes ficticios que se relacionan y viven en un mundo determinado, pero igualmente ficticio. De estos personajes y de este mundo se podría decir lo mismo que dice el sentido de la coletilla que aparece al finalizar muchas películas: cualquier coincidencia con la realidad es pura

coincidencia. La adopción de esta estrategia para el caso de la psicología permitiría considerar a los datos aportados por el experimento psicológico como si fueran "generadores de una ficción del teórico" (1982b, 166). Esta "ficción" no es otra que la teoría que los integra y desde la que se interpretan razonadamente en su variedad y conexión sistemática. En este sentido, Dennett habla de los datos del experimento psicológico como del mundo heterofenomenológico del sujeto de experimentación y del psicólogo como del heterofenomenólogo que enlaza teóricamente esos datos en su conexión sistemática (1982b, 166).

Imaginemos ahora que la novela que el crítico literario quiere interpretar está escrita en estilo autobiográfico. El lector de una novela de este género puede leerla de principio a fin sin pensar en absoluto que el referente de todos los términos en primera persona que aparecen a lo largo de sus páginas es el propio autor del libro. Esto sucede especialmente si el lector tiene por costumbre saltarse las introducciones que en ocasiones acompañan estos libros en las que, quizás, se cuenta que tal pasaje o tal otro está directamente inspirado en no sé que historia real en la que se vio envuelto el autor. Se puede leer la novela *como si* todo lo que se cuenta en ella fuera una *pura ficción*. De modo semejante, si al observar alguno de los datos recogidos, el psicólogo lee, por ejemplo, "yo veo ahora la mancha más pequeña", o "a mi me parece que ahora es algo más grande", no estará necesariamente obligado a considerar que el referente de los términos en primera persona que aparecen en esas oraciones es un sujeto consciente, si por sujeto consciente, repito una vez más, se entiende algo perfectamente caracterizable en desconexión absoluta respecto de las teorías psicológicas con las que contamos para explicar la conexión de esos datos lingüísticos con otros de naturaleza completamente heterogénea. De una forma más directa se podría decir que el mundo heterofenomenológico del sujeto de experimentación no tiene por qué estar inmediata y

necesariamente atado a un yo o a un alma. Esta cuestión se puede posponer siempre que se quiera sin merma alguna para la teoría psicológica.

El punto de vista particular del sujeto es una abstracción objetivamente extrapolable acerca de la cual se puede decir mucho independientemente de si ese punto de vista, como podríamos decir, está o no está *habitado*" (Dennett 1982b, 166).

El último paso de la comparación entre el texto literario y su interpretación y el texto psicológico y la suya lo constituye la cuestión de cuáles son realmente los referentes de los datos de los que dispone el psicólogo a modo de texto. Haciendo *pendant* con el caso de la crítica literaria, no faltan casos dentro de esta área de la hermeneútica en los que se considera que personajes o sucesos descritos en la novela no son en realidad personajes y sucesos puramente ficticios. Antes bien, existen teorías interpretativas dentro de la investigación literaria, especialmente aquéllas de inclinación psicoanalítica, para las que todavía cabe preguntarse con pleno sentido si tal o cual persona de la vida real no fue de hecho el contrapunto o arquetipo que en todo momento inspiró al autor cuando éste concibió a uno u otro personaje, o si tal o cual impresión no fue decisiva para la narración de este o aquel capítulo. Aunque el autor proteste insistentemente y proclame que nada de lo que los críticos dicen tiene pies ni cabeza, no por ese motivo dejarán éstos de insistir en que esas interpretaciones son, cuanto menos, verosímiles, si no completamente ciertas en algunos casos. El autor, como todas las demás personas, está sometido a la ceguera que le impone toda su vida emocional no reconocida y que le hace expresar de un modo más o menos metafórico lo que no puede decirse abiertamente. Sus medios expresivos están, por así decir, esencialmente limitados por los contenidos que es capaz de reconocer. Por lo demás, sea cual sea el valor de esta clase de teorías para la

interpretación de la obra literaria como obra de arte, lo que ellas pondrían de manifiesto muy a las claras es que hay motivaciones profundas en el propio autor que viven ocultas para él, pero que son, sin embargo, el origen de sus expresiones literarias más acabadas.

Ahora es preciso volver la mirada al caso del texto psicológico. ¿De qué puede pensar el psicólogo que habla realmente el sujeto de experimentación cuando aquél tiene delante el mundo heterofenomenológico de éste? ¿Cuáles son los referentes de los datos con los que cuenta el psicólogo? En manos del psicólogo está "usurpar la autoridad del autor y reinterpretar los textos para proporcionar declaraciones de sucesos y procesos "profundos" o, en cualquier caso, inaccesibles que suceden dentro de la cabeza de los sujetos" (Dennett 1982b, 168). ¿No estará el sujeto hablando realmente acerca de estos sucesos y procesos cuando da cuenta de los contenidos que él considera como dados en su experiencia consciente? Finalmente, comenta Dennett, es concebible pensar que el sujeto sólo pueda tener recursos expresivos suficientes para comunicar el texto o aquella parte del texto, que se refiere a los contenidos a los que él tiene acceso consciente, pero que el sujeto está enteramente bloqueado en cuanto a sus capacidades para acceder a aquellos otros estados y procesos a los que haría referencia una reinterpretación adecuada del texto global de que dispone el psicólogo, la reinterpretación adecuada que debería proveer la teoría psicológica.

& 6. Rorty y el pragmatismo.

En el comentario que hace Rorty a propósito de estas ideas fundamentales de Dennett, se insiste en que, desde un punto de vista estrictamente metodológico, estudiar la conciencia es como estudiar cualquier otra cosa que se ponga a tiro y que, por lo tanto, no hay en absoluto

nada especial en la conciencia de lo que cupiera temer un alejamiento progresivo y fatal cada vez que se intentara estudiar empíricamente (Rorty 1982b, 183). En realidad y según insiste Rorty en su nota crítica al artículo de Dennett, si los métodos de la hermenéutica pueden aplicarse con provecho para esclarecer la labor psicológica en su totalidad ello se debe principalmente a que, en el fondo, los métodos de la hermenéutica y los métodos de la psicología empírica, así como, en verdad, los de cualquier otra ciencia natural, son el mismo método, nada más que aplicado a distintos objetos. Unos y otros forman parte de una actitud básica idéntica.

Son diferentes aplicaciones de la doctrina de que ni los textos, ni las personas, ni ninguna otra cosa tienen esencias, o interiores, o naturalezas o significados verdaderos. (Rorty 1982b, 184).

Ahora resulta que todos somos parientes próximos. Heidegger lo era de Mach y Gadamer lo es de Dennett. Todos somos, en algún sentido, "pragmatistas" y operacionalistas.

Una vez que nos hemos asentado cómodamente en la actitud de establecer comparaciones útiles entre distintos ámbitos del saber humano para iluminar su familiaridad, Rorty tiene ahora su turno de palabra. Quien considere que la conciencia tiene una *quidditas* un τὸ τί ἦν εἶναι, una esencia que se escapa irremediabilmente a los métodos de investigación empírica se pone a sí mismo en una situación tan difícil como lo sería aquélla en la que se encontraría quien quisiera defender hoy en día una posición substancialista, al modo aristotélico, para explicar el movimiento local de los cuerpos. Para explicar el movimiento local no hace falta relacionarlo con la noción metafísica de sustancia y definirlo en función de ella y de las nociones de acto y potencia que vienen inmediatamente a continuación. La revolución cultural del Renacimiento llevada a cabo en gran medida por Galileo y Descartes, habría mostrado suficientemente que

aquel procedimiento es innecesario cuando nuestro fin es explicar el movimiento local de los cuerpos. Todos los filósofos que se sitúen en la estela del aristotelismo en este respecto, como Leibniz o Bergson lo hicieron posteriormente, se hacen objeto de esta misma clase de crítica. Pues bien, la psicología cognitiva y la reflexión metateórica sobre ella, harían lo mismo hoy en día por el estudio de la conciencia que hizo antaño la floreciente y revolucionaria filosofía natural renacentista por la investigación del movimiento local. La conciencia también se puede estudiar empíricamente sin temer dejar ningún resto fuera. Si los aristotélicos criticaban a los cultivadores de la nueva ciencia de la naturaleza que sus métodos y supuestos dejaban fuera de la discusión del movimiento local de los cuerpos las nociones de sustancia, potencia, acto y otras semejantes, otro tanto estarían haciendo hoy día los que critican los métodos y supuestos con los que la psicología cognitiva se enfrenta al estudio de la conciencia. Sin embargo, la adopción y el crecimiento de una conciencia histórica suficientemente amplia nos dice que no hay en realidad nada que temer. Si los aristotélicos tenían aquellos temores, tanto peor para los aristotélicos. Si alguien está preocupado porque cree que se corre el peligro de renunciar a algo fundamental cuando se estudia empíricamente la conciencia, tanto peor para él.

Por otra parte, Rorty ha sabido señalar muy bien las consecuencias culturales que tendría aceptar mayoritariamente estas tesis en relación a la conciencia. Desde luego, estas consecuencias no serían menores que las que tuvo en su día la revolución galileana. Estaríamos, de este modo, en medio de una de esas revoluciones que un autor de principios de siglo podría calificar de "formidables", pero de la cual no seríamos del todo conscientes. Nos falta perspectiva histórica. Lo que resultaría evidente, en cualquier caso, es que cuando se avanza por este camino, nuestra cultura occidental se estaría volviendo cada vez más y más antiaristotélica, y cada vez más y

más pragmática, tanto en ontología como en epistemología de la conciencia, o en lo que sea que estemos investigando (Rorty 1982b, 187). Rorty es consciente de que sólo sostener insistentemente la tesis cartesiana de la subjetividad como el último lugar seguro al que cabe volverse con la esperanza de alcanzar un resto de conocimiento absoluto puede mitigar en parte esta irresistible e imparable propensión histórica hacia el pragmatismo y el instrumentalismo en todas las áreas del conocimiento, incluida la del conocimiento de la subjetividad. Pero es evidente que para Rorty este último refugio antipragmatista caerá finalmente, si es que no ha caído ya, ante la insistente presión histórica de los hechos y las teorías. Esta es su fe filosófica racional. En cualquier caso, la relación condicional que quiere extraer es clara: cuanto menos subjetivismo cartesiano y menos aristotelismo esencialista, más instrumentalismo y más pragmatismo (Rorty 1982b, 187).

& 7. Los hechos presionan de nuevo.

Un buen ejemplo vale más que mil palabras para convencer a los cabezotas subjetivistas trasnochados. Al menos, eso se supone. Aquí va. El ejemplo sirve muy bien para aclarar la estrategia heterofenomenológica delineada anteriormente. Una vez más, no me extenderé demasiado en los detalles más técnicos y escurridizos, aunque sí lo suficiente como para comprender el punto que se está discutiendo. Advierto que el ejemplo, aunque completamente acorde en su sentido con el que propone Dennett, no es el mismo que el suyo (Dennett 1982, 169). Dennett habla de la integración ideal de Shakey, el robot desarrollado por Bertrand Raphael para SRI International, con SHRDLU el programa revolucionario de inteligencia artificial diseñado por Terry Winograd. Ahora que la investigación en inteligencia artificial y en psicología pone tanto énfasis en el desarrollo de sistemas inteligentes integrados y en la formulación de teorías unificadas de

la cognición (Newell 1990), es posible reconstruir el ejemplo de Dennett de una manera que no sea tan "ficticia"³.

El ejemplo concreto que ofreceré me lo ha proporcionado una visita que hice a la División de Ingeniería de Sistemas y Automática (DISAM) de la Universidad Politécnica de Madrid. Hace unos años se desarrolló en ese Departamento un sistema flexible de producción llamado "DISAM/2". En su primera versión, el sistema tenía la misión de detectar la calidad de las piezas que venían por una cinta transportadora y hacer con ellas una serie de operaciones. Las piezas podían tener dos formas diferentes y ser transportadas sin seguir un orden, una distancia y una orientación fija y determinada. La misión de DISAM/2 era, en primer lugar, detectar y reconocer esas piezas. Una vez hecho esto un brazo mecánico se orientaba en la posición adecuada para recoger la pieza reconocida, llevarla hasta una máquina herramienta y practicarle allí un taladro. Finalmente, una vez mecanizadas las piezas, el brazo mecánico volvía a colocarlas en la cinta transportadora en una posición determinada. Las piezas defectuosas eran también reconocidas y almacenadas fuera de la cadena flexible de producción.

³ En este sentido, Allen Newell, de la universidad Carnegie-Mellon y John E. Laird de la universidad de Michigan, entre otros muchos, están trabajando fecundamente de unos años a esta parte en SOAR. Este sistema inteligente está concebido principalmente como una arquitectura cognitiva para los procesos simbólicos humanos que sea capaz de explicar cómo es posible la inteligencia y el comportamiento inteligente en el mundo (Newell 1990, 158). SOAR tiene fundamentalmente tres módulos: un módulo lingüístico (NL-SOAR), un módulo que le permite coger cosas de su ambiente próximo mediante un brazo mecánico (Robo-SOAR) y un sistema móvil que depende, a su vez, de un módulo de visión artificial que funciona autónomamente y que sirve para localizar los objetos que se encuentran en su entorno espacial próximo (Hero-SOAR). Según Laird será posible finalmente coordinar adecuadamente estos tres módulos en uno solo de tal forma que se podrá dar a SOAR ciertas órdenes en inglés para que SOAR las responda también en inglés, al mismo tiempo que realiza las demás tareas asignadas.

Los problemas con que se enfrentan aquí los investigadores en inteligencia artificial son los típicos con los que tropieza cualquiera que quiera tener éxito en las tareas de coordinación de los distintos módulos, aparte de los que ofrece cada módulo en particular. Sin embargo, operacionalmente, es decir, en tanto que arquitectura cognitiva que funcione realmente de una forma coordinada en un sistema material, SOAR no está completamente desarrollado (Newell 1990, 194).

Para explicar de qué manera es posible que DISAM/2 haga todo lo que es capaz de hacer, aunque esto sea todavía relativamente limitado, es importante delinear muy genéricamente algunos detalles técnicos. DISAM/2 está provisto de una cámara de TV lineal cuya función es captar las señales luminosas de su entorno próximo y convertirlas en señales eléctricas. Posteriormente, las tarjetas de adquisición y de procesamiento de imágenes, que forman parte del ordenador central, se encargan de recibir las señales analógicas que transmiten las cámaras, convertirlas en señales digitales e introducirlas en la memoria. Como se sabe, estas tarjetas realizan un primer procesamiento de las imágenes a bajo nivel. Gracias a la cámara y a las tarjetas de procesamiento de imágenes es posible transmitir cierta información relativamente muy poco interpretada desde el mundo físico a la memoria del computador⁴.

Dos de las funciones computacionales más importantes de todo sistema de visión artificial son las de *cuantificación* y *muestreo*. La función de cuantificación consiste en la conversión de la amplitud de la señal analógica en distintos niveles de gris o niveles de intensidad. Estos niveles pueden ir desde el cero, que es el nivel que corresponde a un objeto negro o no iluminado, hasta otro nivel superior máximo que correspondería a un objeto blanco o muy iluminado. Entre uno y otro, estarían todos los demás niveles de intensidad de luces y sombras. Por lo que respecta a la función de muestreo ésta consiste básicamente en la transformación de la imagen—más exactamente, de las dos dimensiones espaciales de la señal analógica—en una matriz que se alberga en la memoria de computador. Intuitivamente, esta matriz sería una especie de damero en el que cada

⁴ Todas las aclaraciones y especificaciones que siguen a continuación acerca del funcionamiento, de los distintos componentes y de las operaciones computacionales de los sistemas de visión artificial están tomadas del manual no impreso de José María Sebastián y Zúñiga Introducción a la visión artificial. Madrid: División de Ingeniería de Sistemas y Automática (DISAM) de la Universidad Politécnica de Madrid.

casilla o celdilla está definida unívocamente por su nivel de intensidad lumínica o de gris, y por su posición. Cada una de estas casillas se conoce con el nombre técnico de *pixel*.

Una vez que se ha formado la matriz de *pixels*, el sistema efectúa sobre dicha matriz una serie de operaciones cuya finalidad es la de modificar la imagen ya digitalizada para mejorar alguna de sus características en algún aspecto relevante. Estas nuevas operaciones computacionales sobre la imagen se conocen con el nombre de "preprocesado". Dichas operaciones son de fundamental importancia para que el sistema pueda reconocer posteriormente el objeto, dado que los procesos computacionales de bajo nivel de muestreo, cuantificación y transmisión producen un deterioro notable en la calidad de la imagen. Este deterioro se conoce como "ruido".

Existen diversas operaciones de preprocesado que se pueden realizar con la imagen para eliminar su ruido en uno u otro sentido. Las más usuales son tres: promediado del entorno de vecindad, mediana del entorno de vecindad y promedio de la imagen. Todas estas operaciones tienen sus ventajas y sus inconvenientes. Brevemente, las dos primeras operaciones funcionan de la siguiente manera. Cada *pixel* de la matriz original—cada rectángulo en el damero—está definido por dos coordenadas $f(x,y)$. Sobre cada *pixel* se define una función que transforma la imagen original y genera para ella un entorno de vecindad. Así, por ejemplo, el valor del *pixel* $f(x,y)$ en la nueva función que define su entorno de vecindad sería éste:

$$g(x,y) = w_1 f(x-1, y-1) + w_2 f(x-1, y) + w_3 f(x-1, y+1) + w_4 f(x, y-1) + w_5 f(x,y) + w_6 f(x, y+1) + w_7 f(x+1, y-1) + w_8 f(x+1,y) + w_9 f(x+1, y+1).$$

La reducción del ruido por medio de la técnica de promediado del entorno de vecindad consiste en generar "una nueva imagen $g(x,y)$ cuya

intensidad para cada punto (x,y) se obtiene promediando los valores de intensidad de los *pixels* de f incluidos en el entorno de vecindad" (Sebastian y Zúñiga, 6). Con esta operación se consigue reducir el ruido pero a costa de desdibujar los contornos y otros detalles de forma del objeto.

El análisis de la imagen utilizando el otro sistema que consiste en hallar la mediana del entorno de vecindad genera una nueva imagen $g(x,y)$ a base de hallar la mediana de un entorno de vecindad para cada punto (x,y) de la imagen original. Con este método se eliminarán los puntos de intensidad muy distinta con respecto a los de sus vecinos. En cualquier caso, como dice muy intuitivamente Dennett, todas estas operaciones tienen la función principal de separar el grano de la paja (1982b, 170).

Una vez preparada la imagen de esa manera, el sistema de visión artificial se adentra en las operaciones fundamentales de segmentado de la imagen y en las de descripción y reconocimiento. Mediante las técnicas de segmentado se consigue una partición de la imagen en regiones significativas. Existen diferentes técnicas de segmentado. Entre ellas, por ejemplo, está la de detección de bordes y fronteras. Este método de segmentación, basado en las operaciones de preprocesado, trata de localizar las diferencias de una determinada característica de la imagen en dos regiones que indiquen la existencia de un borde. Normalmente, estas operaciones de detección deben ser completadas con otras operaciones de segmentado basadas en procedimientos heurísticos, dado que, por ejemplo, la iluminación del objeto introduce discontinuidades que no siempre se corresponden con un borde del objeto, discontinuidades que el sistema precisa reconocer.

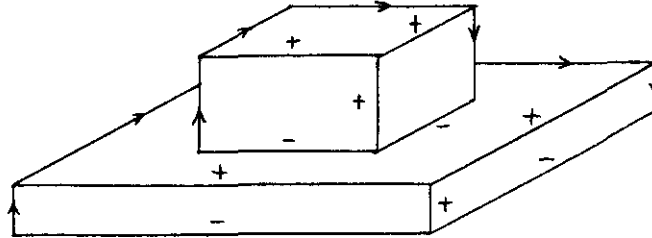
Finalmente, el sistema debe proceder a la descripción de las imágenes y al reconocimiento de los objetos en el mundo. La descripción del objeto consiste en abstraer características de un objeto para su reconocimiento posterior. Estos procedimientos de descripción deben tener la propiedad de ser todo lo independientes que se pueda respecto del tamaño, o de la

localización, o de la orientación del objeto. Un objeto de lados rectangulares, por ejemplo, debe ser descrito por el sistema de tal forma que pueda reconocerlo idealmente esté donde esté en el entorno próximo y sea cual sea su orientación como ocurre de hecho con DISAM/2.

Una vez hecha la descripción del objeto, el paso decisivo de todo el sistema de visión artificial es el del reconocimiento del objeto, esto es, el de la identificación de los objetos existentes en el mundo que constituye el universo de objetos sobre los que opera el sistema. Obviamente, el reconocimiento más interesante en este sentido es el de los cuerpos tridimensionales. Fundamentalmente, hay dos métodos de reconocimiento de objetos, los métodos de decisión teórica y los métodos estructurales entre los que se incluye especialmente los métodos sintácticos para el reconocimiento de las formas o esqueletos de los objetos. La estrategia que rige estos métodos sintácticos consiste en definir un diccionario de uniones permitidas entre vértices para los objetos tridimensionales y en producir un procedimiento de decisión que permita saber, por una parte, cuando nos encontramos ante una sintaxis de aristas permitida en el espacio tridimensional y que genere, por otra parte, un reconocimiento efectivo de la especie del objeto que el sistema tiene delante. En definitiva, este procedimiento debe distinguir un cubo de lo que no lo es. Esta labor se facilita porque el diccionario de uniones permitidas para poliedros sólidos es finito. Sólo un total de doce uniones se permite en cada caso. El sistema empareja estas posibilidades válidas, como si fueran prototipos estructurales que ya posee previamente el sistema, con las aristas que forman las superficies de las imágenes que se han extraído del objeto después de que se han realizado todas las operaciones previamente especificadas.

El etiquetado de las uniones válidas que permite la generación del diccionario se realiza ateniéndose a estas convenciones. En la figura adjunta se ha etiquetado como "+" todas las aristas convexas que resultan de las

intersecciones de dos superficies que son caras de un mismo sólido. Las aristas etiquetadas como "-" son todas aquellas que forman líneas cóncavas como resultado de la intersección de dos superficies de dos sólidos distintos. Las aristas con la etiqueta ">" son todos los bordes de las superficies que ocultan a otras superficies. La especificación relevante en este último caso es que la superficie que oculta a otra superficie es la que se halla a la derecha de la línea mirando en la dirección en la que apunta la flecha. La superficie que no se ve se encuentra a la izquierda de la dirección señalada por la flecha.



Cuando el sistema a realizado el etiquetado de todas las uniones de acuerdo con estas convenciones es cuando está en condiciones de utilizarlas como índices para averiguar la especie geométrica a la que pertenecen los solidos tridimensionales que se encuentran en su entorno espacial próximo.

En el caso concreto de DISAM/2, el sistema de visión identifica de una forma flexible la posición y el tipo de las piezas que vienen por la cinta siempre que éstas no estén superpuestas. Esta información guía posteriormente la orientación y la trayectoria coordinada del brazo mecánico que le permita recoger la pieza en cuestión.

Por supuesto, el programa de visión artificial es relativamente rígido y "monotemático". Cualquier otro objeto para el cual no exista patrones de comparación integrados en el programa no podrá ser detectado por el "sistema perceptivo" de DISAM/2. ¡Qué lástima! Sin embargo, una ojeada a las distintas capacidades del sistema de visión artificial es vital para cerrar el conjunto de puntos de vista de Dennett acerca de la conciencia y de cómo podríamos estudiarla empíricamente. Las nociones de acceso elaboradas anteriormente se ajustan perfectamente a la labor de analizar aquellas capacidades.

En un principio caben tres clases de respuestas posibles a la pregunta: ¿cómo sabe DISAM/2 lo que es un cubo y lo que no es? La primera "contestación" que DISAM/2 podría ofrecer, si se le añadiese quizás un módulo lingüístico, sería muy, muy larga y casi sin sentido para la mayoría de todos nosotros. En ella DISAM/2 sólo hablaría de las múltiples operaciones computacionales que lleva a cabo cuando analiza la imagen, la purifica hasta reducirla a sus contornos y la compara con los patrones que posee; de cómo digitaliza la imagen, la cuantifica, realiza un análisis del entorno de vecindad, la segmenta y la describe y reconoce. En una segunda respuesta diría simplemente que él traza los contornos de sombras y luces, dibuja la imagen resultante y la compara con sus patrones. Así sabe lo que es un cubo y lo que no lo es. En este segundo caso no hace falta mencionar para nada los procesos computacionales que subyacen a estas operaciones de comparación. En la tercera respuesta todo lo que sería capaz de decir DISAM/2 al respecto es que las cosas simplemente le parecen unas veces cúbicas o otras veces no, pero que no tiene ni idea de cómo llega a saber esta clase de cosas y a distinguirlas (Dennett 1982b, 172). El coge las piezas que son cubos y las pone en una parte y las que no lo son en otra. Una vez más, esto último se logra hacer sin postular

ningún ojo interior para DISAM/2 aparte de la maquinaria y de los procesos computacionales que se han mencionado.

La tesis fuerte de Dennett es que, en realidad, las tres respuestas serían igualmente válidas para el caso, pero que la respuesta que sea capaz de ofrecer DISAM/2 depende esencialmente de cuál sea la profundidad del acceso con que sus diseñadores le doten en relación a las distintas operaciones que realiza para distinguir los cubos de otras formas geométricas. En concreto, de cuál sea el grado de acceso computacional que sus capacidades expresivas tengan a las descripciones relevantes en cada uno de esos niveles de análisis de las operaciones realizadas por el sistema de visión artificial. Si las capacidades expresivas de DISAM/2 tuvieran acceso computacional y público a las operaciones computacionales del nivel inferior, seguramente tardaría muchísimo en decirnos lo que hace si quisiera explicarnos con todo detalle su comportamiento discriminativo. Esto, aparte de aburrido, sería bastante improductivo y poco práctico. No; DISAM/2 está diseñado para tener exclusivamente acceso expresivo al nivel superior de descripción.

Las conclusiones que Dennett está dispuesto a extraer de esta clase de consideraciones están bien claras. Si la explicación del comportamiento inteligente de máquinas como DISAM/2 puede servir en algún sentido como modelo para explicar ciertas conductas inteligentes o intencionales humanas básicas--y hay muchas razones para pensar que sí que lo es a pesar de las dificultades que jalonan una empresa de este estilo--, entonces, en el caso humano, existiría también una barrera infranqueable con la que se tropezarían todos los sujetos de experimentación para explicar lo que ellos están realmente haciendo en la situación experimental. Ellos simplemente darían descripciones fenomenológicas. Unas veces las cosas les parecen así y otras así. Esto no significa, por supuesto, que mientan deliberadamente. Realmente a ellos les parece que las cosas son así y así.

Los sujetos, por consiguiente, expresan realmente lo que experimentan conscientemente. Se trata simplemente de que se dejan atrás y "enmascaran"-diríamos que irremediabilmente-todos los acontecimientos y procesos internos que explicarían idealmente su conducta observable, verbal y no verbal. Y son estos procesos los que trataría de encontrar y explicar el psicólogo cognitivo. Los sujetos son, en este sentido, autores de ficción en diversos grados, productores de textos psicológicos, de mundos objetivos y heterofenomenológicos que están al alcance del psicólogo y que éste debe interpretar adecuadamente sin dejarse engatusar completamente por lo que los sujetos dicen y hacen. De este mundo heterofenomenológico comenta Dennett que es un punto de vista exterior de la conciencia del sujeto (1982b, 174). Pero es que no hay ninguno "interior".

Contra este acercamiento al problema de la conciencia cabría revolverse y argumentar inmediatamente que precisamente porque es un acercamiento "externo", es por lo que se deja fuera de él lo esencial de la conciencia, su carácter subjetivo. A esta clase de contraargumentaciones sólo cabe contestar que, en efecto, en el desarrollo actual de la ciencia existen todavía diferencias importantes entre la conciencia de los sujetos en todo su esplendor y el mundo objetivo de datos empíricos que constituyen el texto heterofenomenológico de que dispone el psicólogo. Sin embargo, es una forma de pensar congénita a la idea de la ciencia creer que en la asíntota ambos mundos, el heterofenomenológico y el autofenomenológico del sujeto convergirán definitivamente (Dennett 1982b, 174). Además, es claro que es el punto de vista del sujeto el que orienta y determina los resultados del experimento. El psicólogo, si es un buen psicólogo, no corregirá la plana a lo que los sujetos dicen acerca de sus experiencias; confía en ellos plenamente. Es más, si el psicólogo diseña bien el experimento, no debería dejar traslucir en ningún momento cuáles son sus intenciones teóricas para no influir, o influir mínimamente, en las respuestas

de los sujetos. En este sentido, "*el sujeto tiene la última palabra*" (Dennett 1982b, 177). Lo que puede decir, lo dice y todo el mundo cree en él. En lo que no tiene ni mucho menos la última palabra es a la hora de zanjar la cuestión acerca de qué habría que decir racionalmente sobre lo que experimenta. Acerca de la memoria, de la percepción, de la atención, de la resolución de problemas, cabe decir todavía mucho más de lo que los sujetos pueden decir y describir ateniéndose a sus experiencias conscientes y a los contenidos de estas experiencias. Por este motivo tiene sentido hacer psicología empírica, egología empírica. Finalmente, ciencia cognitiva.

2. INTENCIONALIDAD SIN CONCIENCIA: BRENTANO, SEARLE Y LAS CIENCIAS COGNITIVAS.

& 1. Ciencias cognitivas e intencionalidad.

Había que decirlo decididamente y sin ningún temor. El ambiente no era muy propicio y podían echarle alguna que otra reprimenda. Todos le encarecerían que se olvidara del asunto con gesto comprensivo y condescendiente. Para expresarlo de una forma tan tradicional como tradicional era su tesis: agua pasada no mueve molino. Nadie, o muy pocos, estaría dispuesto a buscar en la dirección en la que él creía que había que hacerlo. Bastante trabajo y tiempo--casi medio siglo de investigación filosófica--había costado desprenderse de las tesis del conductismo lógico como la única alternativa viable en filosofía de la mente que fuera compatible con una imagen materialista del mundo, como para que ahora venga nadie a decir que ese primer paso de gigante es sólo el principio y que, encima, nos hemos olvidado de lo principal.

Hemos recuperado completamente el vocabulario mentalista. Ahora hablamos confiadamente de estados mentales, de contenidos y significados mentales, de intencionalidad, de sistemas de representación mentales. Hemos habilitado, bajo el nombre de "innatismo", un cierto sentido del concepto filosófico tradicional de lo a priori. Y por si esto fuera poco, no hemos dejado de ser materialistas, es decir, hemos elaborado una imagen de la mente que se adecua a las exigencias de la ciencia (psicológica) contemporánea. ¿Qué más se puede querer? ¿Para qué enredar las cosas con quejas vagas como, por ejemplo, que nos hemos olvidado de la conciencia,

de la subjetividad, del problema del tiempo inmanente y de la corriente de conciencia, de la necesaria vinculación de todo estado mental o intencional con un yo y de otros problemas filosóficos de esta catadura? Para decirlo sin ningún rodeo, estamos muy contentos de haber podido disociar para siempre el problema de la intencionalidad de nuestros estados mentales del problema de la conciencia. El lenguaje del pensamiento de Fodor y sus allegados, la gramática filosófica de Chomsky, el funcionalismo de todo género, el computacionalismo de la inteligencia artificial, e incluso la filosofía del lenguaje en general, no necesitan hablar en absoluto de la conciencia, aunque muchos autores que cultivan estas áreas del conocimiento estén dispuestos a hablar confiadamente de estados intencionales. Muy probablemente, todos los procesos mentales innatos y computacionales transcurren completamente inaccesibles a la conciencia.

En medio de este estado de cosas de la investigación en filosofía de la mente, John R. Searle ha alzado tímidamente la voz para decir precisamente que nos hemos olvidado de la conciencia en el camino de las largas, penosas y variadas cuestiones que rodean los análisis en torno al problema de la intencionalidad de los estados mentales¹. De acuerdo con este filósofo, no podría haber sido de otra manera.

Los presupuestos y métodos de la filosofía contemporánea están mal equipados para tratar la *subjetividad* de los estados conscientes. Una aversión semejante a aceptar las consecuencias de la subjetividad ontológica aflige generalmente también a la psicología y a las ciencias cognitivas (Searle 1989b, 194).

¹ Esta manera de proceder en la que se propone separar por completo, o considerar como un caso especial, la teoría de la conciencia y de los contenidos conscientes de la teoría de la intencionalidad más general ha sido denominada por McGinn como "estrategia de aislamiento" (McGinn 1991, 32-33). Para lograr esta finalidad, el modo de operar de esta estrategia consiste en hacer girar el foco del interés teórico hacia el *contenido subpersonal* que parece ser propio de las máquinas y de los procesos inconscientes que tienen lugar en el sistema nervioso. El contenido subpersonal podría ser explicado sin apelar a la conciencia.

De acuerdo con dos de las teorías de la mente más en boga en los últimos veinte años--la teoría representacional de Fodor y la teoría sintáctica de Stich--, las manipulaciones sintácticas de los símbolos mentales que constituyen los procesos mentales causalmente responsables de alguno de los comportamientos de los organismos son, en su gran mayoría, fenómenos inconscientes. Los partidarios de la "mente computacional" no necesitan de la conciencia para dar cuenta de la intencionalidad de los estados mentales. Por el contrario, tradicionalmente, los defensores de la "mente fenomenológica" piensan unidas de una forma necesaria la intencionalidad y la conciencia. Searle se decanta en favor de este segundo punto de vista.

Cualquier estado intencional es, bien de hecho, o potencialmente, un estado intencional consciente y, por esa razón, la ciencia cognitiva no puede evitar el estudio de la conciencia. La atribución de cualquier fenómeno intencional a un sistema, sea "computacional" o de otro tipo, depende de la aceptación previa de la noción corriente de mente, de la mente "fenomenológica" consciente. (Searle 1989b, 194).

Searle se queja en las primeras líneas de su artículo de que no hay o, al menos, es muy extraño encontrar dentro de la tradición analítica en filosofía de la mente, ningún intento serio y con pretensión de sistematicidad de construir una teoría acerca de la estructura general de los estados mentales intencionales conscientes (Searle 1989b, 193). Aunque en esta ocasión Searle no ofrece ningún rudimento de descripción de dicha estructura, es posible intentar su reconstrucción en sus líneas más generales atendiendo a las propiedades lógicas fundamentales de los estados mentales intencionales tal y como se describen en su libro más sistemático sobre estas cuestiones: *Intentionality*.

& 2. La teoría de la intencionalidad de Searle.

(a) Estados mentales y estados intencionales.

Lo primero de todo, conviene aclarar qué entiende exactamente Searle por Intencionalidad. No cabe ninguna sorpresa respecto de lo que tiene que decir sobre este asunto. La Intencionalidad es aquella característica de un gran número de estados y procesos mentales gracias a la cual esos estados y procesos están dirigidos a o son acerca de objetos y estados de cosas en el mundo (Searle 1983, 1). La puntualización introducida en esta primera definición del concepto de Intencionalidad por la que se nos hace saber que no todos los estados y procesos mentales son intencionales tiene su justificación en la observación de que muchos estados mentales que pertenecen a la esfera de los sentimientos—como, por ejemplo, ciertos estados de nerviosismo, de ansiedad, de depresión, de júbilo o de euforia—no son estados intencionales, aunque sí mentales. Sobre ellos, al menos en un gran número de casos, no cabe contestar de una forma determinada a las preguntas ¿de qué estás jubiloso o deprimido?, ¿de qué te alegras? Según Searle, simplemente se está jubiloso o deprimido sin estarlo, al parecer, ¡de nada! Todas las propiedades descriptivas de esa especie de estados mentales conscientes se agotan en ser precisamente eso: un estado consciente de júbilo o de depresión. Esto no quiere decir que de otras formas de ansiedad o de júbilo no quepa contestar de una manera determinada a las preguntas anteriormente formuladas. Esta alternativa constituiría la modalidad intencional de aquellos fenómenos de la esfera del sentimiento. Pero en los casos en los que no somos capaces de hacerlo, Searle piensa que estamos ante el caso de un estado mental consciente que no es intencional.

Por el contrario, de los estados mentales propiamente intencionales como, por ejemplo, las creencias y los deseos, siempre cabría contestar a

las preguntas ¿qué crees?, ¿qué deseas? Sería absurdo por mi parte decir que creo o que deseo sin creer o desear *nada*. Subrayo esta palabra porque Searle parece pensar que mientras que no tiene ningún sentido decir que creo o deseo pero que mi estado de creencia o de deseo no lo es acerca de nada, sí que tiene sentido decir que estoy jubiloso, o nervioso o deprimido--en sus modalidades supuestamente no intencionales--pero acerca de nada.

Desde luego conviene saber si las modalidades supuestamente no intencionales de los estados mentales sentimentales conscientes lo son verdaderamente o sólo relativamente a una cierta limitación de nuestra capacidad descriptiva en el momento en el que estamos viviendo intensamente esos estados. La verdadera situación, según entiendo, sería más bien que siempre me es posible decir con pleno sentido que estoy deprimido o jubiloso de o por algo, aunque ahora mismo no sea capaz de formular muy claramente de qué lo estoy. Podría ocurrir que mi estado de depresión o de júbilo me "llenara" tan completamente que, por decirlo así, no encuentre apropiado (o no fuera capaz de) reflexionar sobre su objeto. La inteligencia, a veces, tiene quizá sus limitaciones para describir el corazón del hombre.

Pero por lo que respecta exclusivamente a mi incapacidad para contestar a la pregunta "¿qué ...?" esta situación no es muy distinta de aquella otra en la que, por ejemplo, sé que quiero saciar mi sed un día muy caluroso de verano, pero ante el puesto de helados no puedo discernir muy bien cuál quiero. Es decir, que si el hecho de no poder contestar determinadamente a la pregunta formulada con el pronombre interrogativo "qué" cuando quiero indagar cuál es el objeto o estado de cosas al que se halla dirigido algunos estados de depresión o de júbilo, fuera una razón suficiente para declarar que aquellos estados no son intencionales, los

mismos indicios se podrían encontrar en algunas especies de estados intencionales que Searle considera incondicionalmente intencionales.

De ser cierto este paralelismo, entonces no parece que la fuerza del único argumento que aquí muestra Searle para establecer que hay estados mentales que no son intencionales sea muy convincente. Si es cierta esta afirmación, entonces los estados mentales son intencionales sin excepción alguna y no cabe decir que muchos estados mentales conscientes--los estados de depresión o de júbilo "indeterminados"--no son intencionales.

(b) La estructura lógica de los estados intencionales conscientes.

Una vez concluida esta digresión, cuya conclusión provisional es importante retener por motivos que más adelante se harán patentes, conviene recordar que el interés primario que guía estas líneas es describir las propiedades lógicas más básicas de los estados mentales intencionales conscientes.

Piensa Searle que, por razones pedagógicas, la mejor forma de analizar la estructura lógica de los estados mentales intencionales es compararla con la estructura de los actos de habla. El resultado de esta comparación es la posibilidad de contar con una descripción de los estados mentales intencionales según estas directrices, por lo demás, tan conocidas.

Decir estado mental intencional es lo mismo que decir contenido representativo bajo un cierto modo psicológico. A su vez, distintos estados mentales intencionales pueden tener muy diversos modos psicológicos pero el mismo contenido representativo. Los modos psicológicos que expresamos mediante los verbos de creencia, de juicio, de deseo y otros semejantes, pueden tener el mismo contenido representativo. Puedo creer que la estatua del *Angel caído* está en el Retiro, pero también juzgar, desear u odiar que la estatua del *Angel caído* esté en el Retiro.

El modo psicológico característico del estado mental determina la dirección de ajuste (direction of fit) entre su contenido representativo y el mundo. Básicamente hay dos direcciones de ajuste, a saber: la dirección de ajuste mente-mundo y la dirección de ajuste mundo-mente. Si se toma como base una teoría muy ingenua acerca de las relaciones entre los contenidos representativos y "el mundo", se puede decir que las creencias y los juicios son modos psicológicos cuyo contenido representativo se supone que tiene que conformarse o coincidir con el mundo—o con lo que Searle llamará más adelante sus condiciones de satisfacción—si es que ha de cumplirse. Si creo o juzgo que la estatua del *Angel caído* está en el Retiro, pero creo o juzgo erróneamente, entonces es el contenido representativo de mi creencia y de mi juicio el que tiene que cambiar, y no el mundo. De aquí su dirección de ajuste característica mundo-mente. Si, por el contrario, deseo u odio que la estatua del *Angel caído* esté en el Retiro, es la situación espacial de la estatua del *Angel caído* la que debería acomodarse a mi deseo o a mi estado de odio y no a la inversa, si es que mi deseo ha de realizarse o mi odio calmarse².

De todos aquellos estados mentales cuyos modos psicológicos tienen una dirección de ajuste, se puede decir que su contenido representativo representa sus condiciones de satisfacción bajo un cierto aspecto. La condición de satisfacción de mi creencia de que la estatua del *Angel caído* está en el Retiro, es que la estatua del *Angel caído* esté en el Retiro, y

² Searle distingue también una tercera clase de modos psicológicos que no tienen ninguna dirección de ajuste aunque se pueda decir de ellos que son apropiados o inapropiados dependiendo de las circunstancias. Por ejemplo, si me alegro de que mi mejor amigo haya hecho bien el examen, mi alegría está fundada, al menos, en dos modos psicológicos con direcciones de ajuste características. En primer lugar, está basada en mi creencia de que mi amigo ha hecho bien el examen y, en segundo lugar, en que deseaba que hiciera bien éste. Si mi creencia es falsa no tiene mucho sentido que diga que me alegro de que mi amigo haya hecho bien el examen. Simplemente no lo ha hecho y, por tanto, mi alegría es inapropiada. Por otra parte, si mi deseo no es sincero, tampoco tiene sentido alguno que diga que me alegro de ello. También en este caso mi alegría sería inapropiada (Searle 1983, 8).

no, por ejemplo, que la única estatua que hay en Madrid dedicada al diablo esté en el Retiro (aun suponiendo que la estatua del *Angel caído* sea idéntica con la única estatua que hay en Madrid dedicada al diablo).

Searle es muy cauto y demasiado ambiguo cuando utiliza el término representación en este contexto.

Decir que una creencia es una representación es simplemente decir que tiene un contenido proposicional, que su contenido proposicional determina un conjunto de condiciones de satisfacción bajo un cierto aspecto, que su modo psicológico determina una dirección de ajuste para su contenido proposicional tal y como todas estas nociones—contenido proposicional, dirección de ajuste, etc.—se explican en la teoría de los actos de habla. En verdad, respecto de lo que he dicho hasta ahora, podría estar dispensado de utilizar los términos "representación" y "representar" en favor de estas otras nociones, dado que no hay nada ontológico acerca de mi uso del término "representación". Es sólo una abreviatura de esta constelación de nociones lógicas tomadas en préstamo de la teoría de los actos de habla. (Searle 1983, 12).

En relación con el uso del vocabulario técnico que emplea Searle, es importante tener en cuenta que hablar de contenido proposicional o de contenido representacional o de contenido intencional de un estado mental, son tres formas de denominar a lo mismo.

Todo estado intencional consiste en un *contenido Intencional* y en un *modo psicológico*. En donde ese contenido es una proposición completa y en donde hay una dirección de ajuste, el contenido Intencional determina las *condiciones de satisfacción*. Las condiciones de satisfacción son aquellas condiciones que, tal y como el contenido Intencional las determina, se deben dar si es que se tiene que satisfacer el estado. Por esta razón, *la especificación* del contenido es ya una *especificación* de las condiciones de satisfacción. (Searle 1983, 12).

Con todo, aún quedan por explicar otras dos nociones que son claves en la teoría de Searle: la de objeto intencional y la de relación entre el modo psicológico y el contenido intencional.

Respecto de la aclaración de la primera, la tesis de Searle es que el objeto intencional de un estado mental es aquello a lo que el estado mental está dirigido (Searle 1984b, 16). Pero aquello a lo que el estado intencional está dirigido son sus condiciones de satisfacción tal y como están determinadas por el contenido intencional de dicho estado mental. Un estado mental intencional no está dirigido propiamente a su contenido proposicional o intencional y sí hacia su objeto intencional. El objeto intencional es lo que satisface el estado intencional. Por supuesto, la función de satisfacción que tiene el objeto intencional sólo la puede cumplir cuando satisface, a su vez, el contenido intencional del estado intencional en cuestión. En virtud del contenido intencional los estados mentales se refieren a o representan sus condiciones de satisfacción bajo un cierto aspecto. Cuando las condiciones de satisfacción son el caso o existen, entonces el estado intencional tiene un objeto intencional. No lo tendrá en el caso contrario.

Según mi teoría, es crucial distinguir entre el *contenido* de una creencia (por ejemplo, una proposición) y los *objetos* de una creencia (por ejemplo, los objetos ordinarios) (Searle 1983, 17).

Por lo que respecta a la segunda cuestión, Searle afirma que en modo alguno hay que pensar la estructura lógica de los estados mentales intencionales como si ésta consistiera en una relación binaria entre el modo psicológico y el contenido representativo, o entre un creyente y una proposición. La razón en la que fundamenta esta tesis es que el contenido representativo o la proposición no es el *objeto* del estado mental, sino su *contenido* (Searle 1983, 18). La idea de Searle es que las creencias son

idénticas con los contenidos representativos o proposiciones en tanto que tales contenidos representativos aparecen precisamente en el modo psicológico de la creencia. No hay ninguna "distancia" intencional en la que se pueda introducir, como una cuña en la veta de un árbol, una relación--de la clase que sea--entre el modo psicológico del estado mental de creencia de un sujeto y el contenido de ese modo psicológico. El único "topos" en el que cabría introducir una relación sería entre el estado intencional--entendido como un todo indivisible compuesto por el modo psicológico y su contenido intencional--y las condiciones de satisfacción que son representadas en ese contenido intencional. "Sólo recuerda ... que es posible que haya el estado intencional sin que haya nada que de hecho lo satisfaga" (Searle 1983, 19). Es decir, que un estado intencional, en tanto que estado intencional, sería la unidad sin fisura alguna entre un modo psicológico y su contenido intencional, unidad que determinaría unas condiciones de satisfacción bajo un cierto aspecto en virtud de la "porción referencial" de su contenido intencional (Searle 1984b, 17). Entre el contenido intencional y sus condiciones de satisfacción se da una relación de representación cuando aquellas condiciones de satisfacción son el caso. Cuando no se dan las condiciones de satisfacción, sólo hay una determinación de cuál sería el estado de cosas que satisfaría el contenido intencional del estado mental. Así, pues, hay que distinguir de una manera precisa entre las funciones de determinación y de representación asignadas al contenido intencional, y sus condiciones de satisfacción. La determinación de las condiciones de satisfacción es una función esencial del contenido intencional; una función que, por así decir, no puede dejar de tener sin dejar de ser un contenido intencional. Por el contrario, se podría decir que la función de representación del contenido intencional le viene dada extrínsecamente por la circunstancia de que sus condiciones de satisfacción sean el caso.

(c). **Intencionalidad intrínseca y conciencia.**

Piensa Searle que la posibilidad misma de hablar de una identidad entre el estado mental y su contenido representativo, en tanto que dado en un cierto modo psicológico, permite solucionar un problema que ha atosigado a la filosofía de la mente desde hace mucho tiempo. El problema es el siguiente. Para que haya representación debe haber un yo, un agente, que use un objeto primario—una imagen, una pintura, unos sonidos—como representación de otro objeto. Por tanto, para que todas las actitudes proposicionales o estados mentales intencionales—como las creencias o los deseos—sean representaciones, debe haber un yo o un agente que use esas creencias y deseos como representaciones. Una vez más, para que sea posible que este agente use las creencias y deseos como representaciones, es preciso que se las represente mediante nuevos estados mentales intencionales, y así sucesivamente. Por lo tanto, se necesita una hilera infinita de *homunculi*, cada uno con sus propios estados intencionales, para que los estados intencionales originales sean representaciones de un cierto estado de cosas. Daniel Dennett ha llamado a este problema "el problema de Hume" (1981, 122).

Lo que Searle tiene que decir sobre este argumento de acuerdo con su teoría, es que una de sus premisas es desafortunada porque descansa en un equívoco. No es verdad que para que haya representación tenga que haber un agente que use un objeto primario como representación de otro; al menos no en todos los casos en los que hablamos de representaciones. Puede resultar que las imágenes, las pinturas o los sonidos que emitimos al hablar sólo representen en tanto que un agente los use como representaciones. En este sentido, es preciso hablar de formas de intencionalidad derivadas. Pero la intencionalidad propia de los estados mentales no es derivada, sino, según la terminología de Searle, intrínseca.

Lo que quiere decir Searle con este término es que no se precisa ninguna forma de "intencionalidad exterior" (outside Intentionality) para que el contenido representacional de un estado mental represente sus condiciones de satisfacción bajo un cierto aspecto (Searle 1983, 22).

Estas afirmaciones sobre la estructura intencional de los estados mentales se deben complementar necesariamente con la siguiente teoría sobre la conciencia:

Por ejemplo, parte de lo que es ser consciente de la creencia de que está lloviendo es ser consciente de que la creencia se satisface si es que está lloviendo y que no se satisface si no llueve...Decir que el agente es consciente de las condiciones de satisfacción de sus creencias y deseos conscientes no quiere decir que tenga que tener estados intencionales de segundo orden acerca de sus estados intencionales de primer orden de creencia y deseo. Más bien, la conciencia de las condiciones de satisfacción *es parte* de la creencia o del deseo consciente, dado que el contenido intencional es *interno* al estado en cuestión (Searle 1983, 22). (Los subrayados son míos).

La mente fenomenológica de la que habla Searle, en contraposición a la mente computacional, es una mente que no sólo tiene creencias y deseos, sino también una mente que se da cuenta de cuáles son las condiciones de satisfacción de sus estados mentales.

Si tal y como sugiere la tesis principal del artículo de Searle que vengo comentando, no es posible una teoría de la intencionalidad sin una teoría de la conciencia, habrá que analizar muy detenidamente la idea de que la conciencia de las condiciones de satisfacción de un estado mental es *una parte* de dicho estado mental, y una parte tal que no implica en modo alguno un estado intencional de segundo orden. Por otra parte, es necesario remontarse aún más allá de lo que lo ha hecho Searle para contestar a la pregunta de qué quiere decir intencionalidad intrínseca por oposición a intencionalidad derivada y cuál es el papel de la conciencia en

la demarcación de esta división, si es que tiene alguno. Mi tesis es que decir intencionalidad intrínseca es lo mismo que decir conciencia, y que una ontología de los estados mentales intencionales es una ontología de la conciencia.

(d) **La taxonomía de los estados mentales.**

El problema con el que se enfrenta Searle surge, a mi modo de ver, por no haber analizado un poco más detalladamente de qué modo se puede decir que la conciencia es una parte de los estados mentales y de qué modo se puede hacer esta afirmación sin que ello implique un estado intencional de segundo orden. De ahora en adelante, llamaré a estas dos preguntas consecutivas, con toda justicia, "el problema de Brentano". La urgencia de estar en posesión de una teoría mínima para resolver el problema de Brentano se ve reforzada por las inconsistencias a las que lleva la teoría de Searle acerca de la clasificación de los estados mentales. Por consiguiente, antes de analizar el problema de Brentano en sus dos vertientes, conviene fijarse en lo que tiene que decir Searle en torno al problema de la clasificación de los estados mentales.

En otros escritos anteriores³ este autor ha señalado ya lo que en este artículo constituye uno de sus temas centrales, a saber, que la división más genérica de los estados mentales que podemos hacer los clasifica en estados mentales conscientes y estados mentales inconscientes. Es decir, un estado mental no deja de ser un estado mental por el hecho de ser inconsciente, por el hecho de que en su descripción el elemento que expresábamos con las palabras "se da cuenta de que..." no aparece por ninguna parte. Para utilizar las metáforas de las que se sirve Searle—bien es

³ John R. Searle, 1984, Minds, Brains and Science (Cambridge, Mass.: Harvard University Press) 16.

cierto que en la conciencia de que son un tanto ingenuas--, los estados mentales son como peces o como muebles. Los peces y los muebles tienen las formas y propiedades que tienen independientemente de si están, respectivamente, fuera del agua donde podemos verlos o en las profundidades del océano; en el cuarto de estar o llenos de polvo en el ático oscuro fuera del alcance de nuestra vista.

¿Cuáles son entonces, según Searle, las propiedades que los estados mentales deben tener para ser precisamente estados mentales y, *a fortiori*, estados mentales intencionales? ¿Cuáles son estas propiedades que no pierden los estados mentales por el hecho de aparecer en una de sus dos modalidades principales: como estados mentales conscientes o como estados mentales inconscientes? Un estado mental debe poseer necesariamente estas dos características:

(i) Ha de tener intencionalidad intrínseca en oposición a la mera intencionalidad derivada o intencionalidad "como si".

(ii) Los estados mentales representan sus condiciones de satisfacción sólo bajo un cierto aspecto que es esencial para comprender cómo el sujeto o agente se representa aquellas condiciones de satisfacción (Searle 1989b, 197).

Según estas dos declaraciones, tanto los estados mentales conscientes como los inconscientes--por ser precisamente estados mentales--poseen intencionalidad intrínseca y representan sus condiciones de satisfacción bajo un cierto aspecto. Por consiguiente, cuando añadimos el adjetivo "intencional" a la expresión "estado mental", no parece que añadamos ningún plus de significado, dado que tanto los estados mentales conscientes como los inconscientes son estados mentales intencionales. Lo que es importante saber ahora es si los adjetivos "consciente" e "inconsciente" funcionan bien modificativamente, bien atributiva o determinativamente,

cuando acompañan a la expresión "estado mental (intencional)"⁴. Es ahora cuando tiene sentido la conclusión anterior de que no hay estados mentales que no sean intencionales. Si los hubiera, quebrarían esta tajante división dual.

Para poder dar una respuesta a la cuestión planteada, es conveniente revisar, antes que nada, el ejemplo que pone Searle de lo que sería un estado mental inconsciente, así como el modo que tiene de describirlo. Para este filósofo, mi creencia, que puede que jamás haya formulado explícitamente, de que la Torre Eiffel está en París (cuando no estoy pensando en dicha creencia), es un estado mental inconsciente. Por ser un estado mental, es intencional y representa sus condiciones de satisfacción bajo un cierto aspecto (Searle 1989b, 196-197)⁵. Es decir, que hay estados mentales intencionales inconscientes que representan sus condiciones de satisfacción cuando están siendo inconscientes. La consecuencia que se sigue de esta afirmación es, según creo, que la intencionalidad intrínseca y el representarse un objeto bajo un cierto aspecto, no parece que tengan nada que ver con la conciencia, dado que los estados intencionales inconscientes *también* exhiben estas dos características. Si es cierto, como acabo de exponer, que esta conclusión se deriva necesariamente de las tesis y ejemplos de Searle, entonces decir "estado mental intencional" no tiene nada que ver con el hecho de que sea consciente o no. Pero esta última es, precisamente, la tesis de los funcionalistas contra la que, teóricamente, quería argumentar Searle en este artículo. Las dos orillas del río en las que quiere estar Searle a la vez son estas dos: si bien es verdad que, por una

⁴ Para la aclaración del sentido en el que utilizo la distinción entre adjetivos en función determinativa y adjetivos en función modificativa, véase (Brentano, *Psychologie* 2:61-62 y Twardowski 12-13).

⁵ La misma clase de ejemplo ofrece Searle en *Intentionality* para ilustrar precisamente la misma tesis. Aquí tampoco añade nada a la descripción y caracterización de los estados de creencia inconscientes en relación a lo que se acaba de exponer (Searle 1983, 2).

parte, no quiere afirmar que la intencionalidad y la conciencia sean lo mismo, tampoco quiere arriesgarse a desligar completamente la intencionalidad y la conciencia como hacen los funcionalistas. Pero la vía intermedia, si es que hay alguna, está poblada de numerosos obstáculos y paradojas, y se corre el peligro de ahogarse en el intento de recorrerla.

& 3. Problemas de la teoría de la intencionalidad de Searle.

Según mi punto de vista, aunque el diagnóstico de este filósofo a propósito del programa de las ciencias cognitivas, según el cual sería posible contar con una teoría de la intencionalidad suficientemente fundamentada sin tener una teoría de la conciencia, es básicamente acertado, sin embargo no ha conseguido ser suficientemente convincente a la hora de dar argumentos para que le creamos. Searle quería sustituir una teoría "objetiva" en tercera persona acerca de la intencionalidad, por una teoría subjetiva en primera persona. Pero la forma de conseguir esta sustitución pasaba por explicar por qué el nexo entre la intencionalidad y la conciencia es mucho más resistente que lo que suponen algunos psicólogos y teóricos cognitivos. La conclusión del párrafo anterior es que si las propiedades (i) y (ii) son determinantes para que algo sea un estado mental y el ejemplo que pone Searle para ilustrar lo que sería un estado mental inconsciente es válido, entonces la vinculación entre la intencionalidad intrínseca y la conciencia se diluye de nuevo.

Tal y como yo lo entiendo, el origen de este fallo en su argumentación es triple. Por una parte está en juego la suposición de Searle, casi convertida en dogma, de que no es posible abandonar el suelo de una metafísica materialista a propósito de la ontología de los estados mentales, tanto conscientes como inconscientes. Por otra parte, se encuentra una vez más el problema de las difíciles relaciones entre la intencionalidad

intrínseca de los estados mentales y la conciencia que tenemos de ellos. En Searle, más que de una relación habría que hablar de una disociación. Finalmente, su noción misma de estado mental inconsciente plantea paradojas que refuerzan la conclusión del punto anterior.

Quizá la forma más clara y sencilla de proceder sea comenzar por esta última cuestión. Searle define los estados mentales inconscientes de la siguiente manera.

Entendemos la noción de un estado mental inconsciente sólo como un posible contenido de conciencia, sólo como la clase de cosa que, aunque no consciente, y quizá imposible de traer a conciencia por varias razones, es, sin embargo, *la clase de cosa* que podría ser o podría haber sido consciente (Searle 1989b, 197).

De esta cita se desprende inmediatamente que un estado mental no sólo tiene un contenido que representa sus condiciones de satisfacción bajo un cierto aspecto, sino que este estado mental también es o puede ser, todo él, un contenido de conciencia. En el fondo, las afirmaciones de Searle se reducen a postular que un estado mental intencional inconsciente tiene un contenido representacional que representa sus condiciones de satisfacción bajo un cierto aspecto, aunque este contenido pueda pasar completamente desapercibido para la conciencia. Sólo bajo ciertas condiciones neurofisiológicas puede este contenido ser traído a conciencia.

Si es importante el texto que he citado en último lugar es porque en él contamos con una pista más para determinar las propiedades de la conciencia. Decir conciencia de las condiciones de satisfacción es decir *parte* del estado mental. Esta parte convierte al todo formado por el modo psicológico y el contenido representacional del estado mental en contenido suyo.

Esta forma de definir los estados mentales inconscientes tiene muchas consecuencias de gran alcance para los problemas que suscita el análisis de

la intencionalidad y su relación con la conciencia. No es únicamente que no haya en la teoría de Searle un apoyo firme en el que pueda descansar la idea de que existe un vínculo interno entre intencionalidad y conciencia, sino que además otras muchas nociones se tornan asimismo problemáticas. Uno de los resultados con los que parece tener un compromiso Searle, es que la caracterización de las nociones de contenido y de representación no necesita en absoluto de la conciencia.

Para no abandonar el terreno de las ideas que han aparecido en la cita anterior, indagaremos qué condiciones se requieren para que "una cosa" sea un posible candidato a contenido de la conciencia y qué quiere decir propiamente esta declaración. La tesis más explícita que se encuentra en Searle a este respecto es que son los conceptos de intencionalidad intrínseca y de "forma aspectual" (aspectual shape) los que hacen que algo sea mental en absoluto y, por consiguiente, que ese algo sea un posible candidato a contenido de conciencia (1989b, 198). Unas líneas más abajo ofrece una nueva pista para que pensemos en la clase de unión que se daría entre la intencionalidad y la conciencia.

La unión, pues, entre *intencionalidad* y *conciencia* se encuentra en la noción de forma aspectual. Para que sea intencional, un estado o proceso debe ser pensable o experimentable, y para ser pensable o experimentable debe tener una forma aspectual bajo la cual sea, al menos en principio, pensable o experimentable (Searle 1989b, 198).

Sin embargo, decir esto es resbalar sobre el duro hielo del problema sin hacer mella en él. Pero es que, además, nada de lo que acaba de decir Searle puede ser cierto si es que se ha de mantener que los estados mentales inconscientes, por ser mentales, son intrínsecamente intencionales y representan sus condiciones de satisfacción bajo un cierto aspecto. Si esto es así, la unión entre intencionalidad y conciencia no puede ser la noción de aspecto o sentido, ya que, como el mismo Searle admite, un estado de

creencia inconsciente como, por ejemplo, aquel que se expresa en el enunciado "la Torre Eiffel está en París", representa también sus condiciones de satisfacción bajo un cierto aspecto. ¿O es que el ejemplo que tanto repite Searle está mal elegido desde un principio?

Cuando enuncié el triple origen de la debilidad de los argumentos de Searle, dije que el problema de las relaciones entre la intencionalidad intrínseca y la conciencia era uno de ellos, y quizá el fundamental. Me parece claro que desde el momento en que este filósofo defiende que los estados mentales inconscientes son intencionales, sólo cabe una disociación radical entre la intencionalidad intrínseca y la conciencia. Se abre en este lugar, pues, la interesantísima perspectiva de analizar hasta qué punto se puede hablar de intencionalidad intrínseca sin hablar de conciencia. En otras palabras, se trataría de averiguar si un estado mental, simplemente en virtud de su contenido, puede ser intrínsecamente intencional sin que lo acompañe necesariamente la conciencia que lo convierte a todo él, de una manera muy peculiar, en contenido suyo. La pregunta, por consiguiente es ésta: ¿No será la conciencia que "acompaña necesariamente a todas mis representaciones", la fuente de toda intencionalidad intrínseca? ¿No será lo mismo decir intencionalidad intrínseca que decir conciencia? Si son éstas las preguntas correctas en este punto, ¿dónde está el error de Searle? Aquí entramos de pleno en la discusión del problema de Brentano.

& 4. El problema de Brentano.

Una hipótesis posible para hallar la debilidad de la teoría de la intencionalidad de Searle es pensar que este autor no ha desarrollado una ontología formal descriptiva suficiente de los estados mentales intencionales. Sólo se puede disociar de una manera tan radical intencionalidad intrínseca y conciencia cuando se ha llegado al convencimiento de que un acto

mental sigue siendo tal, aun cuando no sea el contenido de (un acto de) la conciencia. En otras palabras, para sostener esta tesis acerca de la relación entre la conciencia y la intencionalidad intrínseca—tanto de los estados mentales inconscientes como de los conscientes—, es preciso pensarla según el modelo de los colectivos, con el significado preciso que dio Brentano a este término en su *Psicología* del año 1874.

Brentano considera que los estados mentales intencionales o actos psíquicos y los fenómenos físicos que aquéllos intenden, forman una unidad, una cosa. Los fenómenos psíquicos y sus contenidos—los fenómenos físicos como partes intencionales, que no reales, de los fenómenos psíquicos—, son partes de un fenómeno real unitario (Teile eines einheitlichen wirklichen Seins) (Brentano *Psychologie* 1, 221). El principio que rige la ontología de Brentano en este punto es que es imposible que algo sea una cosa real (ein wirkliches Ding), y que al mismo tiempo sea una multiplicidad de cosas reales; un colectivo (Kollektive) (*Psychologie* 1, 222). Hecha esta aclaración, se destaca que la primera distinción ontológica debe ser la distinción entre cosas reales y colectivos.

Brentano advierte que no debemos dejarnos engañar por el hecho de que en el lenguaje natural nombremos normalmente los colectivos con un término singular para concluir erróneamente de esta circunstancia que los colectivos son cosas singulares. En la vida cotidiana hablamos de un enjambre, de un rebaño o de una ciudad. Pero cuando hablamos así no hablamos de una cosa real, sino de un colectivo de cosas. Sin duda alguna, un problema importante que se plantea inmediatamente es saber qué es una cosa real y, en conexión con él, el de encontrar un ejemplo adecuado de esa peculiarísima categoría ontológica. Un rebaño no es una cosa real porque está formado de ovejas. Estas, a su vez, están formadas de patas, vísceras, etcétera. La dificultad de encontrar los "átomos"—los únicos a los que podríamos llamar con todo derecho "cosas reales"—a partir de los cuales

están formados los colectivos, no es abordada directamente por Brentano, aunque es consciente de ella. Sólo indica que no puede haber colectivos si es que no hay cosas reales (Psychologie 1, 222).

Hecha esta primera aclaración, Brentano establece el segundo principio de su ontología. Aunque ciertamente una cosa real no puede ser un colectivo de cosas, no por ello estamos desautorizados a afirmar que en una cosa no se puedan distinguir multiplicidades. Siguiendo a Aristóteles en este asunto capital, Brentano subraya que unidad no quiere decir necesariamente simplicidad. Dicho brevemente, la unidad real de una cosa puede contener una multiplicidad de partes (Brentano Psychologie 1, 223). El ejemplo que precisamente pone Brentano para ilustrar cómo es posible esta extraña situación ontológica, es el de un acto de representación en el que hay que distinguir como mínimo entre el objeto primario del acto y el objeto secundario—el acto que es objeto de sí mismo—de la percepción interna; de la conciencia, en la terminología de Searle. Esta afirmación requiere, sin duda alguna, una explicación más detallada.

Antes de hacer esta necesaria aclaración, es importante señalar otro punto relativo a la terminología que utiliza Brentano. Con el fin de elegir un término más selectivo y no confundir el uso de "parte" cuando se utiliza para referirse a las partes que constituyen un colectivo—las ovejas son partes del rebaño—, Brentano emplea el término "divisivo" (Divisiv) para designar las partes que se pueden distinguir en un objeto real sin que por ello haya que afirmar que el objeto ha dejado de ser una unidad que se ha pulverizado en numerosas partes independientes. Divisivos son las partes de una unidad que la dejan intacta como tal unidad pero que la convierten en compleja (Brentano Psychologie 1, 223).

Es razonable pensar que Searle no considere la conciencia—la percepción interna, en la terminología de Brentano—como un divisivo de un estado mental intencional. Es por este motivo por el que el adjetivo

"inconsciente", cuando se añade a la expresión "estado mental intencional", funciona determinativamente. Para Brentano únicamente podría funcionar modificativamente. Un estado mental intencional inconsciente no es en absoluto un estado mental intencional, de la misma manera que un hombre muerto, un diamante falso o un falso amigo no son ni hombre, ni diamante, ni amigo algunos. Así, pues, las imágenes más o menos simples que comparan un estado mental intencional inconsciente con la forma que conserva un pez o un mueble independientemente del lugar en donde se encuentren situados, parece que hay que tomárselas muy en serio, si es que Searle tiene razón. Pero cuanto más en serio se tomen, más falsas de tornan.

Si no me equivoco, esta clase de metáforas son las que Brentano rechaza cuando afirma que no puede haber algo así como una conciencia inconsciente (ein unbewusstes Bewusstsein). En esta expresión hay que entender por "conciencia" el estado mental intencional, y por "inconsciente", que ese estado mental no se está apareciendo a la conciencia cuando está representando sus condiciones de satisfacción bajo un cierto aspecto—como diría Searle. En otras palabras, una conciencia inconsciente es un acto mental del cual no es uno consciente. Por consiguiente, no se puede hablar propiamente de intencionalidad si no se habla al mismo tiempo de conciencia, de acuerdo con Brentano. Sin entrar en los detalles de las cuatro pruebas dadas por Brentano para negar la posibilidad de una conciencia inconsciente, quiero, sin embargo, presentar aquí la última de ellas que es, posiblemente, la más importante de todas. Sobre su conexión, por no decir identidad, con lo que Dennett llamaba "el problema de Hume" no cabe duda alguna.

El origen de la teoría de Brentano acerca de la imposibilidad de que pueda darse alguna vez una conciencia inconsciente reside en su interés por refutar el siguiente argumento. Algunos partidarios de la existencia de

estados mentales inconscientes sostenían que si todo estado mental ha de ser consciente, entonces todo acto mental sería un objeto del cual el sujeto es consciente. En tal caso, no sólo debemos tener conciencia del contenido de un estado mental que representa sus condiciones de satisfacción bajo un cierto aspecto, sino también de este estado mental mismo. A su vez, en tanto que este estado mental es consciente--somos conscientes no sólo del sonido, sino también de que lo oímos--, debería ser él también el contenido de otro estado mental, y así *ad infinitum* (Brentano Psychologie 1, 171). Dado que el contenido de un estado mental y el estado mental forman un todo unitario, si existiese esa iteración constante de estados mentales, entonces también debería iterarse cada una de las partes que lo componen y, por tanto, el contenido del "primer" estado mental. Es así que no tenemos conciencia de que el contenido se multiplique indefinidamente, luego tienen que existir estados mentales inconscientes. Si no fuera así, habría que admitir un número infinito de estados mentales implicados en el simple acto de oír (Brentano Psychologie 1, 171). Es decir, que si todo estado mental intencional (todo fenómeno mental) tuviera que ser un objeto de la conciencia, entonces los estados mentales deberían ser infinitamente complejos; lo cual es absurdo.

La solución positiva de Brentano a estas dificultades consiste en conceder a la parte contraria que no puede haber, en ningún caso, un regreso infinito, pero negándole que la única manera posible de evitar el absurdo de un regreso infinito sea postular una conciencia inconsciente. La representación del sonido y la representación de la representación del sonido forman un estado mental unitario. Así, pues, por lo que hace a la ontología más básica defendida en *La Psicología desde el Punto de Vista Empírico*, los fundamentos de la relación intencional son divisivos del acto psíquico considerado como una cosa real unitaria. Hay un estado mental

con dos objetos: sonido y "sí mismo", llamados por Brentano, respectivamente, objetos primario y secundario del acto mental.

La representación del sonido y la representación de la representación del sonido no forman más que un único fenómeno psíquico que dividimos conceptualmente en dos representaciones cuando lo consideramos en su relación con dos objetos diferentes, de los cuales el uno es un fenómeno físico y el otro un fenómeno psíquico. En el mismo fenómeno psíquico en el que se representa el sonido aprehendemos simultáneamente el fenómeno psíquico mismo y, ciertamente, lo aprehendemos de acuerdo con su carácter dual en tanto que tiene en sí al sonido como contenido y en tanto que, al mismo tiempo, está presente a sí mismo como contenido.

Podemos denominar *al sonido el objeto primario* del oír, y al oír *el objeto secundario*. (Brentano Psychologie 1, 179-180).

Sin duda alguna, toda la fuerza del argumento que ofrece Brentano para evitar el absurdo de un regreso infinito de los fenómenos mentales al que lleva irremediablemente el cuarto argumento en favor de la existencia de una conciencia inconsciente, reside en afirmar el "carácter dual" de los fenómenos mentales. No hay regreso infinito en la serie de las representaciones sencillamente porque la serie termina en el segundo miembro (Brentano Psychologie 1, 183). El hilo del pensamiento de Brentano se puede resumir diciendo que el objeto secundario de la conciencia acompaña siempre al objeto primario de la conciencia. De este modo no hay ya más ni regreso infinito, ni conciencia inconsciente.

La conciencia que acompaña a la representación del sonido es una conciencia no tanto de esta representación, como del acto psíquico entero en el cual se representa el sonido y en el que la conciencia misma existe concomitantemente. Aparte del hecho de que se represente el fenómeno físico del sonido, el acto psíquico del oír llega a ser al mismo tiempo su propio objeto y contenido tomado como un todo (Brentano Psychologie 1, 182).

Si Searle acusa a los funcionalistas de no haber ofrecido un análisis más o menos aproximado y completo de un solo estado intencional consciente, creo que Searle tampoco muestra en este artículo ninguno adecuado y, hasta donde yo sé, tampoco lo ha hecho en ninguna otra parte de su importante obra filosófica. Creo que la teoría de Brentano está mejor elaborada y que desde ella se hace más comprensible aquella relación interna entre el contenido intencional y el modo psicológico de la que habla Searle.

& 5. Subjetividad y neurofisiología.

El tercer punto en el que basaba la poca estabilidad de los argumentos de Searle hacía hincapié en su renuencia a pensar siquiera en la posibilidad de que una metafísica no materialista a propósito de la caracterización ontológica, y no ya meramente lógica o descriptiva, de los estados mentales pudiera tener visos de salir adelante en el panorama filosófico contemporáneo.

Estoy contraponiendo "lo neurofisiológico" y "lo mental", pero, por supuesto, según mi punto de vista sobre las relaciones mente/cuerpo, lo mental simplemente es neurofisiológico en un nivel superior (the mental simply is neurophysiological at a higher level.) ... Contrapongo lo mental a lo neurofisiológico de la misma manera a como uno podría contraponer los humanos a los animales sin que ello implicase que la primera clase no está incluida en la segunda. No hay dualismo implícito en el uso que hago de esta contraposición. (Searle 1989b, 209).

De acuerdo con este texto parece que "lo mental" es una *especie* de "lo neurofisiológico" y que, por consiguiente, lo intencional es una especie de lo neurofisiológico, del mismo modo que hombre es una especie del género animal. Para entender debidamente estas líneas de Searle que acabo

de citar, conviene plantear una cuestión un tanto paradójica, a la que este autor se ve abocado como conclusión de su cadena de razonamientos. Expresaba anteriormente la idea de que las buenas intenciones de Searle para enfrentarse con "el miedo a la conciencia en la filosofía analítica contemporánea" (Searle 1989b, 193) no habían conseguido cuajar en un argumento convincente a propósito de las relaciones entre intencionalidad y conciencia. No hay razones para no disociar la intencionalidad de la conciencia si nos atenemos a lo que Searle defiende en este artículo. A pesar de todo, Searle cree que, al menos bajo una cierta interpretación, los estados mentales--tanto conscientes como inconscientes--son "irreductiblemente *subjetivos*" (Searle 1989b, 201). La visión general que ofrece la teoría de Searle es que hay estados mentales; que estos estados mentales pueden ser tanto conscientes como inconscientes; que todos ellos son intrínsecamente intencionales y representan sus condiciones de satisfacción bajo un cierto aspecto; y que todos ellos son, en un cierto sentido, *subjetivos*. La pregunta paradójica a la que hacía mención anteriormente es la siguiente:

¿Cómo podrían ser subjetivos los estados intencionales inconscientes si no hay un sentimiento *subjetivo* hacia ellos, si no hay "qualia", si no hay un como-a-mí-me-parece-sentirlo respecto de ese estado? (Searle 1989b, 201).

La respuesta de Searle es que el modo de existencia de los estados mentales intencionales inconscientes (su ontología), es enteramente neurofisiológica (Searle 1989b, 201). Esta contestación tan decidida tiene como corolario que la ontología de los estados intencionales conscientes habrá de ser necesariamente también neurofisiológica. Los estados intencionales conscientes no son accidentes o modificaciones de una sustancia inextensa, esto es, de una sustancia mental no material subyacente. No podría ser de otra manera si es cierto, como concluía antes, que para

Searle cualquiera de las dos especies en las que se divide el género estado mental son igualmente intencionales. "La única ontología mental es neurofisiológica" (Searle 1989b, 201). En resumen, un estado mental intencional inconsciente es una cosa que posee todas las propiedades necesarias para ser un candidato posible a contenido de conciencia, pero que, por razones que seguramente se podrían explicar en términos enteramente neurofisiológicos, no lo ha llegado a ser todavía. Esta es, según Searle, la principal conclusión de su artículo. Con ella cree haber superado la aparente contradicción que surge cuando se considera que "la ontología de la *intencionalidad inconsciente* es enteramente describible en tercera persona, en términos objetivos neurofisiológicos, pero que, al mismo tiempo, los estados son irreduciblemente subjetivos" (Searle 1989b, 202). Pero, ¿no es esto haber encontrado la cuadratura del círculo?

A estas consideraciones hay que añadir que el peso que pone Searle en la idea de que se puede salvar el carácter subjetivo de los estados mentales inconscientes si se apela a su condición de posibles candidatos a contenidos de conciencia, tiene como trasfondo la tesis de que la ontología en tercera persona de esa clase de estados mentales no consiste en otra cosa más que en propiedades del cerebro que tienen la capacidad de causar un estado mental intencional *consciente* (Searle 1989b, 202). La capacidad que tienen los estados mentales inconscientes de ser virtuales estados mentales conscientes es lo que los hace subjetivos y es, según Searle, la ventaja que tienen respecto de las distintas clases de procesos mentales inconscientes que postulan las teorías contemporáneas en filosofía de la mente. En este último caso, ninguno de estos procesos y estados pertenecerían siquiera a la clase de cosas que son aptas para ser estados mentales conscientes (Searle 1989b, 207). Por ejemplo, no parece que podamos tener la misma clase de acceso consciente a cualquiera de las fórmulas del "Lenguaje del Pensamiento", que la que tenemos a un dolor.

La plausibilidad de que haya algo así como "un sistema de representaciones internas" innato cuyas fórmulas tengan las funciones, entre otras, de ser los objetos de nuestras actitudes proposicionales, de proporcionar las condiciones de posibilidad de una explicación de la capacidad de los seres humanos para entender los lenguajes naturales o para aprender algunos conceptos, depende de largos procesos de razonamiento y no de condiciones neurofisiológicas objetivas del cerebro.

A pesar de que esta crítica de Searle al planteamiento general del problema de la intencionalidad, tal y como aparece en un gran número de teorías contemporáneas de filosofía de la mente, me parece cierta en sus líneas generales, creo que este filósofo no proporciona un fundamento suficiente en el que basar la creencia de que es absolutamente imposible estar en posesión al mismo tiempo de una ciencia objetiva en tercera persona y de una ciencia sobre los procesos y estados mentales formulada en estos términos.

Es literalmente imposible tener ambas cosas, aun cuando (los autores que trabajan en filosofía de la mente) disfracen su error usando un vocabulario que parece mental...pero que ha sido extirpado de cualquier contenido mental. Pueden seguir adelante con esta pretensión porque pueden afirmar que la realidad mental que piensan estar discutiendo es enteramente "inconsciente" (Searle 1989b, 207).

& 6. Resumen.

El objetivo de esta sección ha sido modesto. Su punto de partida lo ha constituido el diagnóstico de la enfermedad que aqueja crónicamente a las teorías de la filosofía de la mente acerca de la intencionalidad. Searle ha dicho que no se puede andar bien si se disocia radicalmente intencionalidad y conciencia. Sin embargo, cuando ha querido poner manos

a la obra para sanar esta fractura, no ha conseguido culminar la operación satisfactoriamente. El motivo fundamental de este fallo reside en que no ha explicado convincentemente la relación compleja que existe entre la intencionalidad intrínseca de los estados mentales y la conciencia que tenemos de ellos. La insatisfactoria teoría de Searle sobre esta cuestión esencial tiene como consecuencia el tener que enfrentarse con "el problema de Brentano" y tomar postura frente a él. Sin embargo, la carencia de una ontología formal descriptiva suficiente de los estados mentales intencionales desbarata desde un principio todo ensayo de solución satisfactoria a este respecto. Las otras dos razones del poco éxito de la operación de Searle, como se ha planteado a lo largo del texto, son, en primer lugar, las paradojas que suscita su noción y caracterización de los estados mentales inconscientes y, en segundo lugar, la defensa de una metafísica materialista a propósito de los estados mentales. El resultado de la conjunción de estos tres problemas es que tenemos que seguir andando con muletas cuando exigimos una teoría de la intencionalidad que no traicione las exigencias de una ontología antidualista a propósito de la naturaleza de los estados mentales, y que tampoco niegue el carácter apremiante de las cuestiones relativas a la intencionalidad intrínseca y a la conciencia como fuente y origen último en el que hay que buscar la intencionalidad intrínseca.

Pylyshyn ha escrito que "probablemente, el segundo rompecabezas más difícil de resolver en la filosofía de la mente (es) el rompecabezas del significado. (El rompecabezas más difícil, en mi opinión, es el de la conciencia, el cual, probablemente, ni siquiera está suficientemente bien definido para cualificarlo como un rompecabezas.)" (Pylyshyn 1985, 24). Este diagnóstico sigue siendo completamente cierto.

CAPITULO 8.

HUSSERL Y LA CRÍTICA AL NATURALISMO EN PSICOLOGIA.

& 1. Fenomenología y filosofía objetivista.

A lo largo de todas las consideraciones precedentes mi objetivo principal ha sido discutir con cierto pormenor algunos puntos básicos de la filosofía de la mente y de la filosofía del lenguaje contemporáneas al hilo de la obra de autores de primera fila que se han ocupado de estos temas. De paso, he intentado también poner un cierto orden y como "sistema" de relaciones entre la variadísima multitud de investigaciones, más o menos dispersas, que tienen en general las miras puestas en el estudio de la intencionalidad, investigaciones que, por otra parte, se consideran de alguna manera, si no profundamente, interrelacionadas.

Si quisiera resumir el esquema general que ha seguido el progreso de la práctica totalidad de las argumentaciones desde el principio hasta el final de este libro, podría hacerlo de una manera sumamente escueta diciendo que todas ellas se basan en dos pasos fundamentales: desde la psicología popular a la ciencia objetiva del ser psíquico. La evocación, casi inmediata, del título del importante libro de Stephen Stich *Desde la psicología popular a la ciencia cognitiva* (1983) no es, ni mucho menos, casual. En este título se recoge acertadamente el núcleo sobre el que gira

la peonza de la actitud *última* y *única* que parece haber sido capaz de hallar la investigación ontológica del ser psíquico sobre la que reflexiona la mayor parte de la filosofía de la psicología actual. En realidad, no cabía otra alternativa cuando la tarea consiste en detallar los compromisos ontológicos de la investigación psicológica y cognitiva ya en marcha, en fundamentar el método y la validez de sus resultados en comparación con otras ciencias especiales, en explotar algunas de sus consecuencias más importantes para la metafísica y en afrontar los problemas más acuciantes con los que parece enfrentarse. Da igual que en el fragor de la discusión racional sobre el problema de la intencionalidad así enfocado se abandone al final la psicología popular y los supuestos sobre los que descansa—como si fueran un prejuicio tan rechazable como los muchos que han atenazado la razón humana durante siglos—, y uno se convierta, consiguientemente, en un materialista eliminativista a ultranza. Da igual que se acepten ambas y se persiga entender en qué consiste el puente de unión entre la psicología popular y la psicología científica tal y como intenté hacer en el capítulo primero siguiendo de cerca en ciertos aspectos el pensamiento de Fodor. Da igual, por otra parte, que se adopte cualquiera de las posiciones ontológicas posibles fundamentales respecto del ser psíquico, sea el monismo anómalo de Davidson, el "biologismo" de Searle o el funcionalismo de la inmensa mayoría. Da igual. La peonza sigue girando y nadie la mueve de su sitio. Pues bien, a esta forma característica de entender y de hacer las cosas la denominó Husserl "objetivismo" o "filosofía objetivista" que describió exactamente de esta manera:

Lo característico del objetivismo (Objektivismus) es que se mueve sobre el suelo del mundo previamente dado ya como obvio por la experiencia y se interroga por su "verdad objetiva", por lo que para él y para todo ser humano racional es incondicionalmente válido, por lo que es en-sí el mundo. (Husserl 1976a, 70).

Por lo demás, si se consiguiera parar o ralentizar un poco el giro vertiginoso de esta peonza quizás alcanzara a leerse inscrito en ella este cartel: "la ortodoxia". A veces las cosas se pueden expresar y resumir muy bien y Haldane ha sintetizado inmejorablemente en qué consiste la tesis de la "ortodoxia" racional en filosofía de la mente cuyo núcleo central es el naturalismo.

Hasta ahora no he dicho exactamente qué es lo que entiendo por naturalismo, aunque mis indicaciones sugieren una noción que es ahora ampliamente utilizada y que lo equipara con el fisicalismo: el punto de vista de que la realidad, o en cualquier caso, el dominio de los particulares concretos, las propiedades y las relaciones es un solo bloque con el mundo investigado por las ciencias naturales. Este punto de vista es ahora un lugar común entre los filósofos de habla inglesa. En realidad, creo que es la ortodoxia, porque, aun cuando en tiempos recientes se ha escrito mucho acerca y en contra del reduccionismo, por ejemplo, en ética y filosofía de la mente, las tesis que sufren el ataque son usualmente conceptuales o semánticas y no se pone en duda tan frecuentemente que la realidad sea en el fondo física. El buen recibimiento del monismo anómalo y de otras versiones del fisicalismo-muestra se debe, en general, a que consigue reconciliar los fenómenos psicológicos con el naturalismo, es decir, con la ontología fisicalista. (Haldane 1989, 307).

La pregunta que plantearían ahora muchos autores, la que es realmente pertinente hacer aquí, es ésta: ¿acaso se puede hacer de verdad otra cosa aparte de naturalismo? ¿Se puede hacer otra clase de ontología del ser psíquico? Si no es el objetivismo, entonces ¿qué resta? Como algunas de las tesis fundamentales del naturalismo y, sobre todo, del naturalismo no reduccionista, se han analizado ya suficientemente en el capítulo segundo, sólo queda por averiguar qué otra alternativa no-naturalista queda disponible en filosofía de la psicología. En este capítulo se analizará una de las posibilidades más importantes y más elaboradas que cabe encontrar, y que pretende superar el sentido naturalista de las

investigaciones en psicología, la de la fenomenología husserliana.

& 2. Actitudes recíprocas: de la fenomenología con la psicología científica, y viceversa.

Presumo, a juzgar por la bibliografía frecuentemente mencionada en multitud de artículos y libros, que en algunos círculos no forma parte del uso del conocimiento histórico general que fue Husserl el filósofo que en nuestro siglo deslindó, con la claridad exigible, la psicología popular y la psicología científica, y el que desarrolló una de las formas posibles de concebir las relaciones entre una y otra. Esta tarea la realizó fundamentalmente en su libro *La Crisis de la Ciencias Europeas y la Fenomenología Transcendental*. Que nadie se asuste por este título. Los he visto más largos y enrevesados. En cualquier caso, su significado se analizará dentro de poco y se hará así completamente inteligible.

Por otra parte, supongo que da ciertamente igual que la distinción entre psicología popular y psicología científica se conozca a través de la famosa distinción de Sellars entre la imagen manifiesta y la imagen científica del hombre--la versión más conocida en aquellos círculos-- o a través de la distinción anterior de Husserl entre mundo de la vida (*Lebenswelt*) y ciencia objetiva. No es, desde luego, acerca de prioridades históricas sobre lo que pretendo discutir aquí. Si me interesa mencionar la obra que Husserl dedicó a la crisis de las ciencias especiales y, en especial, a la crisis de la psicología como ciencia natural, es porque en ella se contiene la tesis más radical que la filosofía fenomenológica transcendental supo encontrar en relación a la fundamentación de la psicología como ciencia objetiva. Es esta tesis la única que quizás podría plantear todo el problema de la fundamentación de la psicología de una forma distinta que la "ortodoxa", o, al menos, darle un tono diferente desde el principio hasta

el fin. Por cierto, que solamente mencionar este punto me pone la carne de gallina... de miedo. Es tal la desconfianza que reina entre numerosos filósofos y psicólogos que se dedican a pensar seriamente sobre los supuestos de la ciencia que ellos están desarrollando que me da ganas de no seguir adelante.

He leído algunos ataques terribles contra los fenomenólogos y los simpatizantes de la fenomenología en general. Pylyshyn, uno de los psicólogos cognitivos más destacados y que más se ha preocupado por los problemas de principio de la ciencia que está cultivando--por cierto que de una manera muy fructífera y acertada--ha dicho estas palabras acerca del libro de Dreyfus *Lo que no pueden hacer los computadores*. Antes de citarlas diré que Dreyfus pasa por ser uno de los filósofos que en el mundo de la filosofía norteamericana más cerca se ha alineado--con mayor o menor fortuna--de la fenomenología. Las palabras de Pylyshyn son éstas:

Sospecho que <el libro de Dreyfus> se puede ignorar ampliamente, excepto por los saboteadores de última hora, los fenomenólogos y los humanistas mal orientados que ven en él la reivindicación de un experto de los prejuicios concernientes a las frías, duras y tecnocráticas ciencias de la mente. (1974, 58).

Antes que nada, quiero advertir que creo sinceramente que gran parte de lo que dice Pylyshyn en esta reseña contra el libro de Dreyfus está justificado y es acertado. Pero exponer esta discusión tampoco es el problema inminente que aquí me preocupa. Lo que me preocupa es que Pylyshyn no cite ni una sola vez la obra de Husserl cuando ataca tan frontalmente a los fenomenólogos. Entiendo, claro está, que en su reseña sobre *este* libro de Dreyfus no tiene por qué hablar de Husserl. Pero tampoco me parece acertado descalificar desde un principio a la fenomenología en su conjunto por *ese* y *otros* libros de fenomenólogos que

Pylyshyn y otros autores puedan tener en mente. Y eso que, repito, gran parte de lo que dice me parece muy adecuado.

Este tipo de actitudes, más o menos justificadas, ha creado sin embargo un estado de opinión muy amplio entre los filósofos de la mente actuales que ha motivado el olvido casi definitivo de la obra de Husserl en relación con la filosofía de la psicología. Yo tampoco voy a tratar de reivindicar aquí la fenomenología *tout court*. Supongo que a lo largo de este libro se habrá podido apreciar que no estoy de acuerdo con Husserl en algunas cuestiones importantes. Mi objetivo, antes bien, es analizar las relaciones entre la fenomenología y la psicología científica tal y como parece que las vio Husserl por si resultara que en ellas hubiera todavía algún elemento teórico digno de rescatar y de someter a discusión crítica. Es este estudio el que no he visto hacer todavía con cierto detalle y amplitud, y, sin embargo, del que depende enteramente que quepa seguir sosteniendo una actitud abiertamente hostil o, cuanto menos, vágamente contraria y despreocupada hacia la fenomenología. La filosofía de la mente contemporánea no lo ha hecho porque está preocupada con otras cuestiones más acuciantes para ella en este momento, y porque la desconfianza reina en el ambiente. Cosa "mística" esto de la fenomenología. Evidentemente, también cabe atribuir esta carencia al hecho de que nadie encuentre el tiempo y las ganas suficientes para hacerlo. Los fenomenólogos, por su parte, o no han prestado en general todavía suficiente atención a lo que supone el giro computacional en psicología y a lo que implica en la historia del pensamiento la aparición en nuestra cultura de las ciencias cognitivas, o consideran de entrada que hay algo esencialmente erróneo en las consecuencias que este enfoque está dispuesto a extraer en relación al estudio de la subjetividad y de la intencionalidad. Así que, bien por unos, bien por otros, la casa está sin barrer.

Con todo, aparte de para someter a crítica, en la medida de lo posible, los posibles prejuicios de uno y otro bando--con riesgo de introducir uno mismo otros nuevos--, me interesa por sí mismo ver qué hay de verdad en las tesis de la fenomenología de Husserl a propósito de los fundamentos de la psicología científica, sobre todo cuando se tiene como trasfondo de la discusión lo que se ha dicho hasta ahora a propósito de la psicología cognitiva. Por supuesto, esta última parte no pretende ser, ni mucho menos, "irenismo" barato, sino un acercamiento al estudio de las cosas mismas.

& 3. Mundo de la vida y ciencia. Relaciones.

En oposición o, mejor dicho, como complemento al esquema de investigación ontológico objetivista que va desde la psicología popular hasta la ciencia objetiva del ser psíquico, el esquema universalísimo que ofrece Husserl en su obra más madura en relación a todas las investigaciones posibles que cabe emprender para hablar de la ontología de la psicología y, en general, de la ontología de cualquier otra ciencia, es tripartito y no dual, o incluso monádico, como sucede también a veces. Exactamente los tres pasos de su marco teórico son éstos:

- (i) Autoevidencias protooriginarias.
- (ii) Autoevidencias del mundo de la vida.
- (iii) Ciencia.

Completar con contenido este esquema es algo que, en parte, se ha hecho ya. Las relaciones entre los puntos (ii) y (iii), para el caso concreto de la psicología, fueron en gran medida el objeto de la discusión del capítulo primero. A estas alturas y antes de abordar los problemas

específicos que plantea las difíciles relaciones del punto (i) con los otros dos-relaciones en las que, sin duda, se guardan las tesis más originales e importantes de la "filosofía de la psicología" de Husserl--, haré una recapitulación para refrescar las ideas allí contenidas al tiempo que las complementaré con algunas matizaciones nuevas.

Husserl atribuye al mundo de la vida, al mundo en el que vivimos, actuamos y del que hablamos cotidianamente antes de hacer ciencia y cuando dejamos de hacerla, una serie de características prioritarias y sorprendentes en relación a la ciencia. Antes del surgimiento de la idea de la ciencia moderna, así como del desarrollo y afianzamiento de sus métodos, y de la obtención de los múltiples resultados objetivos que se han logrado gracias a esos mismos métodos, estaba ya ahí disponible para la humanidad el mundo de la vida. La humanidad tenía detrás de sí una historia de acciones en el mundo y pensamientos sobre el mundo antes de que en la época renacentista europea surgiera el ideal del conocimiento científico. Cuando observo con detenimiento lo que sucede en mi caso, en tanto que individuo de esa humanidad, me doy cuenta de que estoy en una situación muy parecida en relación a la ciencia. También me estaba dado un mundo--y realmente "el mundo"--con determinados contenidos, más o menos especificables, antes de que naciera en mí todo interés teórico científico. Así, pues, en relación a las tareas de la ciencia, el mundo de la vida se presenta, tanto biográfica como colectivamente, como una especie de constante ahistórica y prosigue con este mismo sentido aún después del comienzo histórico de la ciencia y de sus fines propios (Husserl 1976a, 125-126).

Para la argumentación posterior es de primera importancia constatar que para Husserl la nota principal de todos los contenidos que se pueden especificar cuando se reflexiona sobre este mundo de la vida, es la su autoevidencia (Selbsverständlichkeit) (1976a, 124).

El mundo de la vida es el reino de las evidencias originarias.
(Husserl 1976a, 130)

Esta propiedad peculiar de los contenidos del mundo de la vida hace que siempre quepa la posibilidad de volver reflexivamente una y otra vez a este suelo inamovible de evidencias con las que siempre contamos de antemano antes de hacer ciencia. El movimiento característico de la filosofía tardía de Husserl es convertir este suelo básico de evidencias del mundo de la vida en el "fundamento tácito" (verschwiegener Grund) (1976a, 118), en el "constante suelo de validez" (ständiger Geltungsboden) (1976a, 124), en el "olvidado fundamento de sentido" (vergessenes Sinnesfundament), de la tarea de la ciencia así como de sus resultados y métodos (1976a, 48) en una orientación que en seguida se analizará.

Sin embargo, la empresa científica, desde su surgimiento como un tipo de racionalidad humana peculiar, y, en verdad, como la racionalidad por excelencia, ha ido avanzando poco a poco por rutas de pensamiento desacostumbradas para el pensamiento del hombre que vive en este suelo de autoevidencias y no la ha acompañado al unísono en su trayectoria constante. El científico, cuando ejerce como tal y no cuando vive simplemente como hombre en el mundo, es precisamente quien transita estos caminos ajenos al pensamiento de sentido común. La ciencia se convierte, de esta manera, en un *mecanismo* o *instrumento* para el descubrimiento progresivo de verdades no accesibles para el hombre común, para cada uno de nosotros cuando no ejercemos en tanto que científicos. Es así como se ha ido engrosando paulatinamente el tesoro de verdades que se puede alcanzar gracias a los métodos de la ciencia y como se vislumbran otros aún mayores por descubrir en el futuro. Todo consiste en ponerse a la tarea. El problema es que estas verdades teóricas científicas ya a la mano ponen delante de nuestra consideración reflexiva "otro" mundo ajeno y como vuelto de espaldas a esas evidencias del mundo de la vida desde

el que parece que arrancaron aquellas verdades, mundo que, en muchas ocasiones, está además en conflicto con ellas. Aparecen así dos mundos, dos imágenes, posiblemente contrapuestas, surgidas en el desarrollo de la historia de la humanidad.

Desde el punto de vista del objetivismo, al que no es completamente ajeno Husserl, es característico considerar en bloque al mundo de la vida como *mundo y naturaleza precientífico* (Husserl 1976a, 50). De la misma manera, es también característico de la filosofía objetivista considerar que el mundo de la ciencia debería reemplazar *teóricamente* al mundo precientífico; las verdades de la ciencia a las evidencias precientíficas del mundo cotidiano.

La peculiaridad del pensamiento último de Husserl respecto de las relaciones entre vida precientífica y ciencia es considerar que, de alguna manera, las verdades de la ciencia son un "ajustado ropaje de ideas" (ein wohlpassendes Ideenkleid) que ha venido a sustituir y a encubrir, como si se tratara de un "disfraz", al mundo de la vida y a las evidencias en él contenidas y vividas originariamente así por nosotros (1976a, 51). Obviamente, esta sustitución no significa que, por ejemplo, en vez de ver cosas rojas, tengamos que ver ahora longitudes de onda, fórmulas matemáticas o cosas por el estilo. No es que tenga que cambiar la faz de nuestra experiencia cotidiana con el conocimiento de aquellas verdades científicas. Se trata más bien, como apuntaba antes, de una sustitución teórica. Lo que quiere denunciar Husserl al hablar de las verdades de la ciencia como de un "ajustado ropaje de ideas" que enfunda al mundo de la vida que le da sentido, es el peligro en que incurre el objetivismo cuando pretende que "tomemos por ser verdadero lo que es un método" (1976a, 52), que tomemos por ser aquello que no se nos da si no es a través de un método científico, y de un trabajo teórico y empírico largo y riguroso que se ajuste a ese método. Respecto de este pensamiento de

Husserl, de momento no se puede calibrar bien por qué lo hallado por medio del método científico no es el verdadero ser, o cierto dominio del ser, sino algo que "despista" al verdadero pensamiento del ser. En cualquier caso, lo que importa resaltar ahora es el lugar preferente que otorga siempre Husserl al mundo de la vida respecto del mundo de la ciencia.

Las imágenes que utiliza Husserl para hablar de las relaciones entre uno y otro mundo sugieren en todo momento la existencia de una gran bola de nieve que con el paso de la historia se ha ido haciendo cada vez más y más grande, pesada y arrolladora, pero en cuyo centro se encierra un núcleo pétreo modesto sin el cual jamás se hubiera llegado a formar esa enorme bola que llamamos ciencia. Ese núcleo es el mundo de la vida preteórico y el de sus evidencias. Por otra parte, nuestra curiosa posición respecto de esa bola es como si estuviéramos a la vez dentro y fuera de ella. Fuera de ella porque somos capaces de verla, inmensa, y de pensar que si hay bola es porque hay chinita en su interior. Dentro también, porque en nosotros está aposentada esa chinita y rodamos inevitablemente en el interior de la bola y, en cierto sentido, casi al unísono con ella. Desde esta perspectiva, la crítica de Husserl al objetivismo es que no se puede pretender eliminar o sustituir, ni siquiera teóricamente, la piedrecita sin destruir o, al menos, alterar esencialmente el sentido que tiene la bola de nieve.

La teoría objetiva en su sentido lógico... radica, se fundamenta, en el mundo de la vida, en las evidencias originarias pertenecientes a él. Por mor de ese enraizamiento la ciencia objetiva tiene una constante relación de sentido con el mundo en el que vivimos siempre y en el que también vivimos como científicos y, entonces, también en la comunidad general de cocientíficos—así pues, con el mundo de vida universal. (1976a, 132).

Lo curioso de la situación, tal y como la describe Husserl, es que la importancia y la fuerza que ha alcanzado la construcción científica desde la época moderna ha ocasionado que no se vuelva a reflexionar y a poner en claro cuál era el suelo de ideas que estaba en el comienzo antes del surgimiento de toda ciencia. La bola de nieve es tan imponente y mueve con tanta fuerza nuestra visión del mundo natural que no nos deja ver su pequeño centro. Uno de los rasgos más característicos de esta etapa de la obra de Husserl es señalar que no hacer claridad sobre este punto de partida y olvidarse de este fundamento de sentido es el criterio básico para detectar que una ciencia especial está en crisis. La crisis de las ciencias europeas, a la que alude el título del libro de Husserl, es precisamente el estado de olvido constante en el que se encuentran las diferentes ciencias especializadas respecto de este fundamento originario, es dejarse enceguecer por el resplandor de la gran bola de nieve y desatender a su centro esencial (1976a, 10). No obstante, Husserl es cauto y no deja de señalar también que semejante crisis no significa, ni mucho menos, que las ciencias deban dejar de avanzar cada vez más y más, de una manera teórica y práctica, en la prosecución de sus fines, y que deban preocuparse de inmediato por este otro problema relacionado con el esclarecimiento de sus orígenes. Esta es ya una labor del filósofo. Pero sí señala insistentemente que no preocuparse ya más por él supone poner en serio peligro el "sentido de verdad" de las ciencias (Husserl 1976, 10).

Así, pues, y para resumir, el mundo de la vida frente al de la ciencia se presenta, como poco, con todas estas características. Es una realidad presupuesta antes de toda idealización metódica científica; los contenidos de esta realidad son constantes e invariables históricamente; es el olvidado fundamento de sentido de la ciencia natural; es el suelo primitivo de toda vida teórica y práctica y el que le da sentido propiamente en cada estadio de desarrollo.

Por otra parte, Husserl piensa que la explicitación de los contenidos autoevidentes del mundo de la vida constituye de por sí una tarea teórica sistemática en la que es posible encontrar una serie de descripciones que pretenden ser objetivamente válidas. Una cierta ciencia, en un sentido de ciencia diferente del de la ciencia natural objetiva, se ocuparía de investigar estas evidencias originarias. Estas descripciones tienen el sentido de poner en claro ciertas propiedades con las que se nos da de antemano este mundo precientífico que es tan importante respecto del surgimiento de la ciencia, de su sentido y de sus logros teóricos. De esta manera, parece que son tres las tareas de investigación que, como mínimo, debe emprender todo filósofo:

(i) Intentar hacer explícitas todas aquellas "autoevidencias" en las que se está y con las que se cuenta desde siempre; "hacer imagen" lo que es sólo subsuelo de creencias; reflexionar sobre lo que nunca se reflexiona.

(ii) Pensar sobre las verdades de la ciencia, sobre sus objetos, sobre sus principios de pensamiento y sobre sus métodos.

(iii) Meditar sobre la propia naturaleza del pensamiento filosófico y su lugar dentro de este marco aquí esbozado.

Este esquema, común tanto para Husserl como para Sellars, es el que, de alguna manera, se puso en marcha en el capítulo primero para el caso de la psicología. Diré de paso que la diferencia fundamental entre Husserl y Sellars respecto de este tema estriba, precisamente, en el lugar que debe asignarse a la filosofía dentro de él. El complejo y único lugar que le asigna Husserl se estudiará en este capítulo. En el caso de Sellars, es al menos claro que una de las funciones principales de la filosofía es la de analizar todas las posibles relaciones entre (i) y (ii). Pero ambos, Husserl y Sellars, comparten la idea de que presentar y describir los contenidos autoevidentes del mundo de la vida, de la imagen manifiesta, es un campo de investigaciones filosóficas relativamente independiente de todos los demás

que puedan concebirse. Sin embargo, para Sellars, la imagen manifiesta es un *marco de pensamiento conceptual* que se tiene irreflexivamente, con el que se cuenta, con el que no "hacemos" nada, sino desde el que hacemos todo y pensamos originariamente la realidad. Este marco conceptual es un *sistema* que puede llegar a ser descrito y construido correctamente. Pero, en tanto que marco de pensamiento conceptual, puede ser sometido todo él, como cualquier otro, a crítica. Husserl, por el contrario, si bien es cierto que piensa que el mundo de la vida también puede ser sometido a severo análisis, no considera que constituya un sistema de pensamiento que pueda ser criticado con los mismos cánones que cualquier otro sistema de pensamiento, sea éste científico o no, debido, sobre todo, al lugar privilegiado y esencial que ocupa el mundo de la vida respecto de las construcciones teóricas científicas. En este sentido, es claro que Husserl y Ortega y Gasset están alineados en el mismo círculo de pensamiento.

Las creencias constituyen la base de nuestra vida, el terreno sobre que acontece. Porque ellas nos ponen delante lo que para nosotros es la realidad misma. Toda nuestra conducta, incluso la intelectual, depende de cual sea el sistema de nuestras creencias auténticas. En ellas "vivimos", nos movemos y somos. Por lo mismo no solemos tener conciencia expresa de ellas, no las pensamos, sino que actúan latentes, como implicaciones de cuanto expresamente hacemos o pensamos. (Ortega y Gasset 1977, 22).

& 4. Necesidad de una psicología empírica.

Para comenzar, Husserl, en su filosofía más madura, sostuvo que en la labor de esclarecimiento de las relaciones entre el mundo de la vida y la ciencia existía inevitablemente un característico movimiento de vaivén o zigzag. Este movimiento alternativo consiste en que no es posible alcanzar un grado de comprensión suficiente de los orígenes que dan sentido a una ciencia especial--a sus métodos y a sus tareas--si es que no se piensa con

seriedad acerca de los resultados que esa ciencia ha alcanzado en sus diversos estados de desarrollo histórico y, en especial, en el estadio en que esa ciencia se encuentra para cada nueva generación de científicos y pensadores. Sin embargo, no es posible tampoco entender el desarrollo histórico de esa ciencia, "en cuanto desarrollo de sentido" (Sinnesentwicklung) , si no se reflexiona, a su vez, sobre estos mismos orígenes (Husserl 1976a, 59).

Por lo que hace al caso concreto de la psicología, esto es precisamente lo que, de alguna manera, de la manera "objetivista", he intentado hacer a lo largo de los dos primeros capítulos. De la psicología popular, del mundo de la vida, a la psicología cognitiva, y vuelta a empezar. Todo lo que implica esta estratagema se puede resumir adecuadamente en estas palabras. Preguntarse por las posibilidades de principio que rigen la psicología, o la biología, o la física, o cualquiera de las ciencias especiales, sólo es posible una vez que la investigación ha llegado lo suficientemente lejos en la persecución de la idea que guía cada una de esas ciencias. Esto es tanto como decir que el motivo para preguntarse por las posibilidades de principio de la aplicación universal del método científico a un dominio concreto de objetos es el "éxito" de ese mismo método después de que la investigación ha avanzado lo suficiente en ese dominio. Es algo así como si la posibilidad estuviera contenida enteramente en la eficacia.

Sin embargo, y por lo que respecta al fin que me he propuesto, no es enteramente imprescindible detallar al milímetro cómo Husserl dio cumplimiento a esta estrategia de zigzag para el caso de la psicología, cosa, por otra parte, bastante difícil de hacer, dado que, según creo, en ningún momento ofreció un estudio minucioso de todos los obstáculos que podrían hallarse en su prosecución y cumplimiento, siquiera sea para la clase de psicología que estaba más en boga en su época: el estructuralismo de

Wundt y Titchener. Por mi parte, concibo que todas las numerosas controversias en filosofía de la psicología entre las tesis del fisicalismo estricto positivista y las tesis del funcionalismo materialista no reduccionista, tal y como se expusieron en el capítulo segundo, son un intento serio por instalarse en el lugar adecuado para solventar las dificultades con las que se encuentra actualmente la psicología y las ciencias cognitivas en la persecución de su idea. En este sentido, bien podría entenderse que todas aquellas discusiones se encuadran dentro de este mirar alternativamente a los logros de una ciencia en un momento histórico dado y a sus posibilidades de principio. Pero si bien es cierto que puede notarse una cierta carencia en la obra de Husserl a la hora de especificar dónde están y cuáles son estas puertas por las que se ha de pasar en este movimiento característico de "slalom", este filósofo tuvo, entre otros muchos, el mérito innegable de señalar condensadamente las ideas directrices que rigen en todo momento el proyecto de la psicología como ciencia natural.

Dos afirmaciones son el transfondo constante de su pensamiento a propósito de la psicología como ciencia empírica. Estas dos afirmaciones se siguen, precisamente, de la idea de la psicología como ciencia natural. Más exactamente expresado, esas afirmaciones adquieren su significado concreto sobre el hecho de que la psicología "tiene evidentemente el sentido de una ciencia "objetiva" de lo subjetivo" (Husserl 1976a, 129). La primera afirmación aludida es que "la psicología es una ciencia de *hechos*, de *matters of fact* en el sentido de Hume" (Husserl 1950, 8). En segundo lugar, que "la psicología es una ciencia de realidades (Realitäten). Los "fenómenos" de que trata... son sucesos reales, que, en cuanto tales, cuando tienen existencia real, se insertan con los sujetos reales a que pertenecen dentro del orden del mundo uno del espacio y del tiempo o de la *omnitudo realitatis*" (Husserl 1950, 8).

Antes que nada hay que constatar con claridad que Husserl tampoco desarrolló una idea suficientemente precisa acerca de cómo los estados psicológicos conscientes de un sujeto--las *cogitationes*--se podrían insertar en la *omnitudo realitatis*; es decir, cómo esos estados, en tanto que sucesos reales, son, finalmente, nudos cualesquiera en el orden de la causalidad natural universal; cómo los actos de conciencia son pensados y se convierten en sucesos de la naturaleza; bajo qué condiciones sería posible esta inserción; qué aparato conceptual sería necesario mover para dar cuenta de ella (Husserl 1950, 70-71). Que contamos de antemano con esta idea, que la inteligencia es inteligencia encarnada, realizada en un sistema material complejo, parece ser el punto de partida más probable que encontramos cuando nos pensamos a nosotros mismos como organismos psicofísicos que actúan en el mundo. Tal y como se vio en el capítulo primero al tratar acerca de cuáles eran las propiedades más relevantes del ser psíquico, cuando su caracterización se realiza adoptando el punto de vista del sentido común, del mundo de la vida, sedimentado en nuestros usos lingüísticos y formas de pensar cotidianos, dos de las propiedades fundamentales de ese "ser anómalo" son su intencionalidad y su capacidad causal. En aquellos análisis hubo ocasión de resaltar en qué sentido Husserl se mostraba especialmente reacio a hablar de la causalidad de los estados mentales. En su lugar, optaba por una versión mucho más relajada de la relación entre los estados mentales y el orden del mundo--muy próxima, después de todo, a la de Locke--y hablaba simplemente de un "imperar" o un "dominar" del yo sobre su cuerpo vivo (*Leib*). Sin embargo, también traté de mostrar entonces que en la propia filosofía fenomenológica de Husserl había elementos teóricos suficientes para hacer fuerte la crítica que se opone a esta manera más diluida de presentar las cosas. La idea de cinestesia ofrecía el lugar donde el yo se daba la mano con el cuerpo sobre el que "impera", el lugar donde el yo se desmembra en órganos particulares sobre

los que manda (Husserl 1976a, 109). La argumentación corría por el camino de mostrar que no cabe hablar realmente de cinestesias si es que no se habla, a su vez, aunque sea ingenua y acríticamente--sin "metafísica", como apuntaba Husserl--, de los estados intencionales como estados con capacidad causal. A este resultado fue, precisamente, al que se llegó. La noción de cinestesia era la "glándula pineal" de Husserl¹. Por consiguiente, en los primeros pasos ingenuos de la investigación--desde la psicología popular--cabía describir perfectamente el ser psíquico como aquel ser peculiar que exhibe lo mismo propiedades intencionales que causales. Pero aunque esto fuera así, también era cierto que en absoluto se había llegado a saber nada concreto acerca de cómo estas ideas, en su armazón universal, podían rellenarse con algo más que con un proyecto de investigación que desea llevarse a cabo hasta el final. En todo caso, estábamos ante el fundamento inteligible de la psicología científica, pero en absoluto ante su fundamento ontológico. ¿Supuso Husserl, quizás, que el problema de detallar el fundamento ontológico de la psicología, por no ser un problema de estricta fenomenología, no caía directamente en el ámbito que él dibujó para la filosofía? Sin dejar de ser cierto que éste no es un problema de rigurosa fenomenología, no por eso deja de ser un problema filosófico enorme. Al menos tradicionalmente siempre lo ha sido y, evidentemente, sigue siéndolo.

Como acabo de insinuar, Husserl dio por hecho que los estados psicológicos de un sujeto forman parte del orden del mundo. Con lo que no podía contar dentro de su mundo cultural era con una idea que fuera fecunda tanto desde el punto de vista teórico, como empírico, e, incluso también, desde el técnico, acerca de cómo esa "inserción" podría suceder;

¹ Esta discusión, por cierto, pondría de manifiesto, como también se anotó en el capítulo primero, que las autoevidencias del mundo de la vida que quieren captar las primeras descripciones de nuestras formas de pensar precientíficas no se pueden explicitar sin cierta dosis importante de controversia. Estas descripciones se pueden hacer mejor o peor, más o menos refinadamente. Este hecho es una pista para pensar que puede haber descripciones, si no erróneas, sí al menos cargadas de prejuicios. La hermenéutica sabe mucho de esto.

cómo ella podría explicarse realmente. Hoy se proclama que hay razones de peso para sostener que se cuenta con una idea feliz desde la que poder comprender mejor esta exigencia interpretativa y hacer con ella algo más que una brillante especulación preteórica. Quizás Husserl pensara que era una cuestión abierta a la investigación empírica futura el dar con semejante idea. Lo que en ningún caso pareció sopesar suficientemente fue la posibilidad de que contar con una buena teoría empíricamente contrastada acerca de cómo es posible un sujeto en el orden del mundo, supusiese un reto a lo que tiene que decir sobre las relaciones entre fenomenología y psicología científica.

En cualquier caso, del pensamiento global de Husserl se desprende que moverse en zigzag desde la psicología popular a la psicología científica es, independientemente del ejambre de dificultades contenidas en esta estrategia, plantear las cosas sin salir nunca de la tesis general del objetivismo y del naturalismo (Husserl 1976a, 102). No obstante, también debería de quedar clara de dónde surge la necesidad de hacer una psicología empírica: de la propia reflexión acerca del ser psíquico desde los supuestos del mundo de la vida.

& 5. ¿Cómo hacer una psicología empírica?

En el capítulo segundo traté de sintetizar las líneas directrices genéricas que guían la investigación en psicología cognitiva, así como de esclarecer alguno de los supuestos metacientíficos esenciales que permiten construir, con un importante apoyo empírico, la "mente metodológica". De aquellas reflexiones preliminares quiero recordar ahora los tres supuestos básicos que posibilitan la construcción de los experimentos psicológicos que son fundamentales para la elaboración de las teorías psicológicas.

(i) El psicólogo que ha diseñado la situación experimental sabe mucho más acerca de ella--acerca de las propiedades que tiene el mundo en un momento determinado--que el sujeto de experimentación.

(ii) El psicólogo cuenta con la evidencia previa de que el sujeto de experimentación es, precisamente, un ser psicológicamente semejante a él, de tal manera que si los datos recogidos son fiables y la teoría que se diseña para dar cuenta de ellos soporta suficientemente la contrastación con nuevos datos, esa teoría podrá aplicársela el psicólogo a sí mismo en tanto que ser psicofísico.

(iii) El psicólogo utiliza como instrumento heurístico universal de sus investigaciones teórico-empíricas la tesis de que hay correlaciones entre el comportamiento públicamente observable del sujeto de experimentación y los estados y procesos mentales inobservables de dicho sujeto, estados y procesos que, presumiblemente, son las causas de algunos de sus comportamientos.

En seguida resalta que estos tres supuestos están centrados en la relación que mantiene el psicólogo con el sujeto de estudio que es puesto en la situación experimental, al igual que en la relación del psicólogo con ciertos supuestos interpretativos y heurísticos. Entre estos supuestos interpretativos se encuentra, en primer lugar, la instancia intencional a la que alude constantemente Dennett. El resultado de combinar todos estos diversos elementos fue la posibilidad de que el psicólogo interpretase el "texto psicológico" sin necesidad alguna de establecer una conexión necesaria entre la interpretación intencional y la conciencia tal y como es típico de la estrategia de aislamiento. Por el contrario, la interpretación intencional daba pie para explotar las distintas nociones de acceso subpersonal inconsciente con las que, supuestamente, se explicaría cualquier cosa que tengamos, o podamos, decir con sentido acerca de la conciencia sin necesidad de tener que hablar de la "luz interior del yo". De esta manera

se habría alcanzado a solucionar o, por lo menos, a aclarar de una manera naturalista el problema de la conciencia y con él, uno de los escollos fundamentales que el filósofo de orientación metateórica podría solucionar en relación a los problemas de fundamento de la psicología (cognitiva). Pero el psicólogo y el filósofo de orientación metateórica deberían darse cuenta de que su propia conciencia quedaría así también idealmente explicada. Ellos mismos en tanto que "sujetos" de percepciones, de actos atencionales o memorativos, serían igualmente explicables de esta manera. No quiero apelar aquí a ninguna sensibilidad, ni menos aún dañarla, para hacer fuerte un argumento. Sólo pretendo llamar la atención sobre una de las consecuencias que tiene el plantear el estudio de la psicología y el de sus problemas de fundamento de la manera objetivista.

En cualquier caso, lo que más me interesa destacar ahora es que el psicólogo ocupa un puesto privilegiado en todo este complicado asunto. Literalmente, no hay nada en todo lo que se ha dicho hasta ahora en lo que él no tenga una participación esencial, algo respecto de lo cual él no sea como su "origen" y "centro". Si siguiera utilizando la terminología de Dennett, el heterofenomenólogo, el psicólogo que se encuentra con datos empíricos y con teorías formadas de juicios acerca de cuya validez quiere estar seguro, parece arrogarse un lugar tan importante en la propia fundamentación de su ciencia—al menos como sujeto que la juzga de principio a fin—que es difícil pensar y aceptar sin más argumentaciones que él mismo pueda quedar "explicado" en función de las nociones de acceso computacional y público antes desarrolladas. ¿Será quizás todo esto un puro prejuicio emocional? Quizás.

Por lo demás, la discusión previa había dado como resultado parcial el hecho de que el diseño de los experimentos empíricos que sirven como base para la contrastación, rectificación y afinamiento—para la falsabilidad en definitiva—de las teorías psicológicas computacionales supone el punto de

vista heterofenomenológico. El psicólogo es el heterofenomenólogo respecto del sujeto de investigación, respecto de esa oscuridad que nos manda sus heraldos para que los descifremos. Sólo quien se instale en este punto de vista "sorprenderá" a la mente en su funcionamiento real y podrá postular sus propiedades más esenciales bajo los supuestos generales del realismo científico. Por lo demás, este punto de vista "externo", como hubo ocasión de analizar, tenía importantísimas consecuencias, aún difíciles de calibrar con exactitud, en relación al estudio empírico de la conciencia y a su ontología. La cuestión que se plantea ahora es, más bien, esta otra: la adopción inevitable de este punto de vista heterofenomenológico en la psicología empírica, ¿no supone, como condición de su posibilidad, el punto de vista fenomenológico, el punto de vista del psicólogo que lleva a cabo la interpretación del "texto psicológico"? En otras palabras, la perspectiva heterofenomenológica, ¿es ella misma realmente suficiente para dar cuenta de todos los problemas que plantean los métodos y compromisos ontológicos de la psicología cognitiva, o toda ella, a su vez, toma como punto de partida no discutido y pasado constantemente por alto la perspectiva fenomenológica? Si hubiera que contestar afirmativamente a esta segunda parte de la disyunción, ¿cuál sería exactamente ese punto de vista fenomenológico y qué dice todo él a propósito de la actitud que el psicólogo debe adoptar en relación a los problemas de método cruciales y últimos de su ciencia? Y, en cualquier caso, la necesidad de adoptar definitivamente la orientación fenomenológica, ¿querría decir que este punto de vista es anterior en algún sentido al punto de vista heterofenomenológico? ¿Sería, sobre todo, la relación entre ambos una relación de fundamentación (fenomenología) a fundamentado (heterofenomenología)?

En relación a estas preguntas y para comenzar por un sitio especialmente importante, es esencial destacar que hay un elemento en la

filosofía de la psicología de Husserl que está completamente ausente de todas las discusiones actuales acerca de los fundamentos de la psicología cognitiva y, en general, de la psicología como ciencia natural. Este elemento ausente es, sin embargo, el más imprescindible que hay a juicio de Husserl si es que se tiene la intención de bucear hasta el final no sólo dentro del fundamento último de la psicología como ciencia objetiva, sino también de la tarea de la filosofía respecto de todas las demás ciencias especiales. Esta tesis, defendida perseverantemente por Husserl a lo largo de gran parte de su extensa obra filosófica, según diversas modulaciones, dice así:

La fenomenología es la instancia para juzgar las cuestiones metodológicas fundamentales de la psicología. Lo que ella afirma en general tiene el psicólogo que reconocerlo, y dado el caso defenderlo, como condición de posibilidad de toda su metodología ulterior. Lo que pugna con ello distingue al *contrasentido psicológico de principio*, exactamente así como en la esfera física todo lo que pugna con las verdades geométricas y las verdades de la ontología de la naturaleza en general es característico del *contrasentido de principio en la ciencia natural...* Seguro estoy de que en un tiempo no demasiado lejano será bien común la convicción de que la fenomenología (o la psicología eidética) es para la psicología empírica la ciencia metodológicamente fundamental, en el mismo sentido en que las disciplinas matemáticas materiales (por ejemplo la geometría y la foronomía) son fundamentales para la física. (Husserl 1950, 193-194).

Esta tesis, si fuera verdadera, merecería sin duda una atención especial, dado que gran parte de la labor actual de reflexión metateórica sobre la psicología y las ciencias cognitivas vive de espaldas, o cree vivir de espaldas, a la "geometría" fenomenológica. En concreto, merecería inevitablemente la atención de todos aquellos que, interesados en la filosofía de la psicología, buscan ahondar cada vez más en los problemas primeros de fundamentación de esta ciencia. Es mi impresión, sin embargo, que semejante tesis ha sido fruto de un olvido muy peculiar y generalizado, aunque quizás no absoluto, por parte de los filósofos de la mente

contemporáneos. Ya comenté indirectamente antes que un índice externo, que no completamente extrínseco, de semejante olvido se manifiesta en las escasísimas ocasiones en que alguna obra de Husserl aparece citada en las bibliografías de los libros y artículos dedicados a estos temas. Más frecuente, en cambio, es la utilización de la obra de Brentano, que, por otra parte, no pasa en un gran número de casos del recuerdo de que fue él quien situó de lleno el problema de la intencionalidad en el eje de las discusiones de la filosofía de la mente como título genérico para todas sus investigaciones. El análisis riguroso del problema planteado—¿es la fenomenología el fundamento de la psicología como ciencia empírica?—no se instaaura, ciertamente, en este nivel de historia "externa". Son las tesis de Husserl sobre el transfondo de "las cosas mismas" las que deben debatirse pormenorizadamente.

& 6. El sujeto empírico y el sujeto transcendental.

La investigación precientífica del ser psíquico había puesto en peligro cierto acercamiento a las pretensiones fundamentadoras de la fenomenología respecto de la psicología empírica. Me refiero en concreto al hecho de que pensar cómo es posible el sujeto en el orden de la naturaleza dejaba abierta las puertas de par en par a la investigación ontológica natural del psiquismo humano que echa mano de todos los instrumentos y herramientas que están disponibles para la ciencia. Hay realmente necesidad de una psicología empírica. Hay necesidad, pues, de ir más allá de la fenomenología en este ámbito de investigaciones. No es, por consiguiente, la parte de la fenomenología que reflexiona sobre el ser psíquico desde el punto de vista del mundo de la vida la que podría erigirse en la "ciencia metodológicamente fundamental". Incluso cuando se piensa en la manera en que la fenomenología concibe los conceptos psicológicos fundamentales se

hace también patente esta necesidad de seguir avanzando por el sendero del naturalismo.

En relación a este asunto es importante hacer las siguientes consideraciones. Como se ha repetido en múltiples ocasiones, la psicología acepta como punto de partida de sus investigaciones un gran número de conceptos que tienen su origen en el lenguaje psicológico cotidiano. Se hablaba, en este respecto, de una *continuidad categorial* entre la ciencia psicológica y el sentido común. Conceptos como los de memoria, atención, percepción, sensación, lenguaje y un largo etcétera, no son patrimonio exclusivo del lenguaje especializado de la ciencia psicológica, sino que forman también parte importante del lenguaje cotidiano. La actitud usual de la ciencia psicológica en relación a estos términos tan importantes para la contrucción de sus teorías, es la de dar por sentado que existe una precomprensión suficiente de los fenómenos psicológicos a los que ellos apuntan; al fenómeno de la memoria, al de la atención, al de la percepción, etc. La tarea del psicólogo que se ve a sí mismo como un investigador más de la naturaleza, en este caso de la naturaleza psíquica, no es otra que la de formular teorías explicativas de estos fenómenos. Si esta tarea supone la precomprensión de los fenómenos a los que se refieren aquellos términos, precomprensión que se facilita sobre todo gracias, entre otras cosas, al uso de dichos términos en el lenguaje cotidiano, no puede, sin embargo, quedar reducida a investigar cómo utilizamos normalmente esos conceptos, o en qué contextos, ni tampoco a la mera caracterización descriptiva de los fenómenos precomprendidos. Ha de pasar, antes bien, a construir una teoría con la que los términos psicológicos preteóricos se carguen con nuevos contenidos, los auténticamente científicos y técnicamente aprovechables. Lo que dice exactamente Husserl a propósito de los términos psicológicos es que la fenomenología tiene que proporcionar "las normas para el sentido y el contenido científico de los conceptos de cualquier

fenómeno" (Husserl 1965, 46). No habrá verdaderamente ciencia de lo psíquico si es que no se procede previamente a la constitución de estas normas. Pero, ¿a qué equivaldrían estas normas? ¿Quién y con qué derecho podría promurgarlas y sancionarlas? ¿Por qué esas normas habrían de ser aceptadas por el psicólogo? ¿En qué sentido son esas normas no sólo el complemento, sino también lo que haría que la psicología fuera, por fin, ciencia, cosa que Husserl no niega que sea?

Entiendo que esta investigación que Husserl propone podría ser asimilada con cierta facilidad a una investigación preteórica de sentido común de los conceptos psicológicos. Esta investigación podría adoptar una vez más, o bien el carácter de prejuicios, o bien el de ser el fundamento inteligible de la ciencia psicológica realmente constituida. Desde este punto de vista, no hay aquí por qué hablar de normas, y menos aún de normas sin las cuales la psicología no podría alcanzar nunca el estatuto de ciencia, sino de preconcepciones sedimentadas en el lenguaje cotidiano a lo largo de los siglos, aunque sean preconcepciones útiles para los fines de la comunicación ordinaria y para el surgimiento de la idea de esa ciencia. No es desde luego tampoco por este camino por el que el psicólogo se vería obligado a transitar de la psicología a la fenomenología.

A la luz de estas aclaraciones cabría preguntarse, por consiguiente, qué expectativas le queda a la fenomenología en su deseo apasionado de ser el fundamento último de la psicología e, incluso, de todas las demás ciencias. ¿Hay en la filosofía de Husserl otros lugares en los que se pueda encontrar las tesis que ofrezcan los frutos apetecidos? Indudablemente, en una obra tan rica y extensa como la husserliana es posible encontrar estos refugios que, por otra parte, quedarían relativamente inafectados, si no completamente, por las críticas previas. ¿Cuáles son estos refugios?

Antes de intentar exponerlos, no quiero dejar de pasar por alto otro escollo que la actual investigación metateórica del ser psíquico podría

sortear. Me refiero al de la posibilidad de investigar empíricamente la conciencia. Una vez que en el camino de ida se alcanza a esclarecer de qué modo la conciencia se convierte irremediabilmente en uno de los puntos más conflictivos de la reflexión metacientífica acerca de los problemas de fundamento de la psicología cognitiva, no caben demasiadas dudas de que la conciencia es algo que se le da al psicólogo en forma de mundo heterofenomenológico del sujeto de experimentación. Este es un resultado que Husserl aceptaría de buena gana, dado que, en principio, asume que la conciencia humana y, en general, la conciencia animal, puede darse perfectamente en la experiencia psicológica y ser, por consiguiente, objeto de la investigación científica empírica (Husserl 1950, 175). Pero, desde luego, no es tampoco en la crítica de la posibilidad de un estudio empírico de la conciencia donde la fenomenología se descubriría como más fecunda para la psicología, para sus métodos y para sus problemas de fundamento. Por otra parte, en relación a este asunto no considero demasiado arriesgado aceptar que el análisis de la posibilidad de investigar empíricamente la conciencia, así como también el análisis de la manera en que se podría llegar a realizar este estudio, y de los supuestos que están en juego en él, está mucho mejor planteado y desarrollado en reflexiones como las de Dennett. No; si la fenomenología se puede explotar hoy en día de una forma fecunda para los problemas de fundamentación de la psicología como ciencia habrá que atender primariamente a textos tan difíciles de interpretar como éste:

Por otra parte <la conciencia empírica> entra en la fenomenología bajo la modalidad del quedar colocado entre paréntesis el mundo entero con sus individuos psíquicos y sus vivencias psíquicas: todo ello como correlato de la conciencia absoluta. Así, entra en escena la conciencia según diversos modos de apercepción y relaciones, diversos dentro de la fenomenología misma, a saber, en sí misma, una vez como conciencia absoluta, otra, en el correlato, como conciencia psicológica incluso en el mundo natural-con distinto valor

en cierto modo y no obstante sin pérdida de su contenido propio en cuanto conciencia. Son estas relaciones tan difíciles como extraordinariamente importantes... En especial, sólo podrá la psicología, tan enérgicamente ambiciosa en nuestro tiempo, lograr la radical fundamentación que aún le falta, cuando disponga de vastos conocimientos acerca de las señaladas relaciones esenciales. (Husserl 1950, 175-176).

A partir de aquí comienzan un sinfín de dificultades de verdad complejas. Los fenomenólogos se encontrarán ahora como pez en el agua. Los filósofos de orientación metateórica, por el contrario, empezarán a ponerse nerviosos y a torcer el gesto, y querrán volverse cuanto antes a casa. No creo, sin embargo, que los primeros tengan tantos motivos para estar contentos, ni los segundos para sentirse tan crispados.

Pero lo primero es lo primero. ¿En qué consiste esta diferencia entre conciencia absoluta y conciencia psicológica? ¿Cómo y en qué sentido, no obstante, son la misma conciencia? ¿En qué radica, sin embargo, su diferencia de valor? ¿No será que las dificultades en este ámbito provienen del hecho de que la relación, desde un inicio, está tan mal dibujada por el propio Husserl que ni siquiera él mismo la llegó a ver con la claridad que siempre se exigió a sí mismo? A lo mejor es que no hay nada que relacionar.

Desde luego, sólo quien estudie dentro de la tradición fenomenológica podrá encontrar alguna indicación más o menos precisa respecto de estas preguntas y encontrar alicientes suficientes, apoyados y supuestos en muchos casos por la marcha de esta misma tradición, para emprender la investigación necesaria. No parece que en la tradición "analítica" se llegue siquiera a sospechar tan claramente esta diferencia y puedo imaginarme incluso que más de uno huiría espantado con sólo oír hablar de ella. ¿Será porque es completamente absurda? Desde luego, bien podría ser ésta la respuesta. Por consiguiente, la primera tarea será apuntar

antes que nada hacia el sentido de estas afirmaciones. Si no se logra hacer esto con la debida claridad, entonces la fundamentación de la psicología científica adquiriría, quizás, un aspecto muy diferente del que le confiere Husserl. Nada está decidido.

La dicotomía entre el sujeto empírico--la conciencia empírica--que pertenece con pleno derecho al orden natural como una más de sus partes y que, por tanto, está sometido también al juego universal de la causalidad, y el sujeto transcendental--la conciencia absoluta--*que no es una parte del mundo*, recorre la filosofía primera más interesante de Husserl de un extremo a otro (Husserl 1963, 65; 1976a, 81). Qué quiere decir todo esto sólo se hará comprensible después de las explicaciones que siguen.

§ 7. De la reducción transcendental a la conciencia absoluta.

Por desgracia, no hay forma posible de entender cómo llega Husserl a establecer esta dicotomía y en qué consiste ella realmente, si uno no se adentra de lleno, por un camino o por otro, en el problema de la reducción transcendental, que es tanto como decir que no es posible entenderla si no se adentra cada cual, el psicólogo incluido, en sí mismo. Pero, ¿con qué finalidad tendría el psicólogo que preocuparse por esta dicotomía y por entender su sentido? ¿Por qué tendría que realizar, consecuentemente, esta reducción transcendental? ¿Por qué habría de dejar momentáneamente de lado su difícil labor como psicólogo, que tan fecunda puede resultar, y dedicar parte de su tiempo--al menos una vez en la vida--a este otro problema? ¿Qué ventaja o claridad intelectual última iba a extraer del método de la reducción transcendental en relación a los problemas de fundamento de su ciencia? ¿Es acaso el problema de la reducción transcendental algo más que una discusión técnica de escuela

filosófica? ¿Cómo se puede la psicología beneficiar de él? Demasiados interrogantes. Es importante proceder poco a poco.

Hay algo que parece claro en la filosofía de Husserl. Si el psicólogo consigue interesarse de veras por la investigación de lo que pueda haber de verdad en las tesis de Husserl acerca de las relaciones entre la fenomenología y la psicología científica y, por consiguiente, preocuparse alguna vez por el problema de la reducción transcendental, dejará de ejercer como psicólogo, como científico natural, y se convertirá, de pronto, en filósofo, en auténtico filósofo. Curiosamente, sólo mediante esta conversión conseguiría fundamentar radicalmente su actividad teórica como científico. Así, pues, Husserl no ve en el fondo ninguna posibilidad de mediación real entre la labor filosófica y la ciencia psicológica, o, en general, entre la filosofía y la ciencia natural. La investigación metacientífica que asigna Dennett al filósofo interesado en el problema de la intencionalidad, en todas sus ramificaciones, es, desde luego, muy interesante y fructífera desde muchos puntos de vista, pero, por desgracia, no es estrictamente filosófica. De nuevo, aparece la separación de tareas y, en verdad, de una forma realmente tajante.

Para volver al conjunto de los interrogantes planteados, ¿de qué se quiere hablar propiamente cuando se utiliza ese nombre tan extraño de "reducción transcendental"? ¿Qué es, precisamente, aquello tan importante a lo que Husserl quiere aludir con él y que parece ser el paso último y necesario para alcanzar la fundamentación radical de la psicología y del resto de las ciencias?

Hay una discusión entre los autores que están directamente interesados en la fenomenología que versa acerca de cuántas vías diferentes hay que distinguir dentro del método genérico de la reducción trascendental propuesto por Husserl. Estas distinciones son, ciertamente, muy útiles y esclarecedoras. Sin embargo, para alcanzar a comprender el punto de vista

final que adoptó la filosofía primera de Husserl en relación a todas las ciencias y, en concreto, a la psicología como ciencia, no seguiré un tratamiento rigurosamente separado de cada una de estas vías, si no que, más bien, pondré en juego, allí donde sea necesario, elementos propios de cada una de ellas. Como alguno de estos mismos autores reconoce, ni siquiera el propio Husserl llegó a distinguir nítidamente entre todos estos caminos (Kern 1962, 304). Me serviré, por consiguiente, de esta relativa confusión de límites y procuraré explotar las ventajas que tiene mantener esta pequeña dosis de caos.

Creo que para Husserl la puerta de acceso principal a la epojé, el otro nombre que recibe la reducción transcendental, es, en el fondo, única. Ella consiste especialmente en una *actitud* y, concretamente, en un *cambio de actitud* (*Umstellung*) que tiene su origen en una exigencia que el filósofo se impone libremente a sí mismo en el comienzo de su tarea. Quien no adopte, o no esté dispuesto a adoptar por los motivos que sea, esta actitud, no entrará por esta puerta, y si no lo hace, se verá afectado por desventajas teóricas prácticamente insuperables. Entre otras cosas, habrá perdido toda oportunidad de fundamentar realmente la psicología como ciencia, porque, por ejemplo, estará ciego para ver las diferencias y relaciones esenciales que existen entre la conciencia absoluta y la conciencia empírica que tan fundamentales son para aquel fin. El lema en este caso dice así: si no hay reducción transcendental no hay conciencia absoluta. La actitud requerida surge, como comentaba, de una exigencia y, en concreto, de una exigencia que todo filósofo, mediante una resolución peculiar de su voluntad, debe imponerse en el comienzo de su actividad filosófica, y que deberá guiar posteriormente el desarrollo de todas sus investigaciones. Esta exigencia nace de la idea, heredada históricamente con especial fuerza desde Descartes, de la filosofía como ciencia absolutamente fundamentada y carente de supuestos, y de las ciencias especiales como ramas de la filosofía así

concebida. En realidad, el objeto final al que tiende aquella especialísima determinación de la voluntad de la persona que comienza a filosofar y que considera como su tarea primordial, es poseer todos los conocimientos que quepa encontrar en la filosofía primera, entendida, en sus rasgos principales, a la manera cartesiana. La exigencia aludida queda perfectamente resumida en estas palabras: abstenerse radicalmente de aceptar como válidos todos los datos, juicios y teorías con los que contamos de antemano de una forma "acrítica", tanto en la vida cotidiana, como en la vida científica, hasta que su valor como conocimientos no se haya sometido a una crítica radical.

En la formulación de este imperativo teórico surge en seguida la dificultad de justificar en qué sentido y por qué todos los conocimientos y convicciones que parecen estar más a la mano, sean los usuales que barajamos a diario, como los auténticamente científicos, deben de ser puestos en suspenso, al igual que todas las realidades a los que ellos se refieren. Queda también por resolver por qué esta crítica, si no es la filosofía misma, sí que constituye al menos la propedeútica inevitable que da entrada a la filosofía verdadera. En el caso del pensamiento de Descartes se puede hacer aquella justificación con relativa facilidad. Tanto en el caso de la existencia de los objetos que forman parte de nuestro entorno cotidiano, como en el caso de los objetos del pensamiento—de los juicios y teorías—, la posibilidad de concebir el más mínimo motivo de duda hace que inmediatamente intentemos dar un paso más atrás que nos devuelva a un terreno donde no quepa ya duda alguna. Este retroceso se realizará cuantas veces sean necesarias en todos los ámbitos teóricos concebibles hasta dar con conocimientos y juicios de los que sea absolutamente imposible dudar más. Estos conocimientos y juicios son los únicos que deberían formar parte de la ciencia filosófica auténticamente tal, y en los que podría descansar finalmente la voluntad del filósofo que se ha impuesto libremente como objetivo construir esta ciencia radical sin presupuestos. Curiosamente,

justamente en ese momento, es decir, solamente cuando el pensamiento pueda descansar en el suelo inamovible que forman aquellos conocimientos que no están sujetos ya más a dudas, es cuando se habría alcanzado la "fundamentación absoluta de la ciencia", tanto de la ciencia natural, como de la lógica (Husserl 1963, 49).

La forma que propone Husserl en las *Meditaciones Cartesianas* para acceder a este estrato básico de evidencias indubitables es especialmente clara e interesante. En su esquema mínimo consta, según creo, de cuatro pasos fundamentales, a saber:

(i) Análisis de la idea de la ciencia y descubrimiento progresivo de que en esta idea está encerrada implícitamente la noción de evidencia (*Evidenz*).

(ii) Análisis formal de las nociones que cabe encontrar dentro de la idea general de evidencia. En este sentido se distinguirá necesariamente entre evidencias adecuadas, inadecuadas y apodícticas.

(iii) Aplicación de la idea de evidencia apodíctica al modo en que se nos da la existencia del mundo.

(iv) Aplicación de la idea de evidencia apodíctica al modo en que se me da mi propia existencia y al modo en que me doy yo a mí mismo como *sujeto que juzga*.

El análisis del punto (i) lo realiza Husserl en el párrafo cuatro de la primera de las *Meditaciones Cartesianas*. Aunque no me detendré especialmente en su examen, sí que es fundamental dejar establecido que toda ciencia, en virtud de la idea misma que la anima, está formada por un sistema de juicios entre los que debe haber una relación de fundamentación. Esta noción de fundamentación, contenida tempranamente en los *Prolegómenos a la lógica pura*, significa que la verdad de unos juicios debe de estar asentada demostrativamente en la verdad de otros juicios previos cuya verdad también se acepta. Los primeros juicios serían *juicios*

mediatos respecto de los segundos, que serían *juicios inmediatos* y en los que se fundarían precisamente aquellos otros *mediatos*.

Después de hacer esta aclaración, Husserl establece que la posibilidad de la existencia de esta cadena demostrativa requiere apelar a la noción de evidencia. La fundamentación de los juicios *mediatos* en los *inmediatos* no significa otra cosa más que aceptar que la evidencia con la que se me da la verdad de los juicios *mediatos* depende de la evidencia con la que se me da la verdad de los juicios *inmediatos*. En este sentido, no es concebible la idea de fundamentación y, por consiguiente, de ciencia, sin la de evidencia. Es esta noción la que paso a considerar inmediatamente de acuerdo con las directrices husserlianas más básicas.

En un primer acercamiento, hay que entender por evidencia- "evidencia en el sentido más amplio posible" como dice Husserl- el hecho de que tengo "*experiencia* de la existencia o de la esencia de las cosas" (*eine Erfahrung von Seiendem und So-Seiendem*). "*experiencia*", en definitiva, de la existencia de *X*, o de que *X* es *Y* (Husserl 1963, 52). Sin embargo, a lo largo del curso de las experiencias cognoscitivas posibles se hace patente que la evidencia con la que vivo la verdad de un determinado juicio puede ser "más o menos perfecta" (Ibid.). Lo que antes juzgaba que existía, ahora resulta que no existe; lo que me pareció *Y* resulta que en realidad es *Z*. En otras palabras, que el grado de perfección de la evidencia está en función de la corregibilidad de los juicios que formulo acerca del ser o del ser- así de las cosas. Esta clase de correcciones, como puede comprenderse, son comunes tanto para los científicos en el transcurso efectivo de sus investigaciones, como para los legos en su actuación cotidiana. La diferencia fundamental entre unos y otros es que el grado de evidencia que quiere alcanzar la ciencia es el máximo posible, es decir, que la verdad de los juicios a los que aspiran las teorías científicas no debería estar sometida ya más a aquel proceso constante de corrección. Estas

verdades han de ser válidas "de una vez por todas y para todos" (Husserl 1963, 53). Sólo así la ciencia se erige en auténtica ciencia. Sólo así las verdades son realmente verdades. Sin embargo, únicamente en la asíntota de la investigación científica se podría hallar esta noción de evidencia perfecta, de verdad para siempre. En este sentido, la idea de evidencia máximamente perfecta y su correlativa de verdad absoluta, son ideas reguladoras de todos los esfuerzos teóricos y cognoscitivos y, en primer lugar, de los esfuerzos de la ciencia.

Si la ciencia, como a la postre debe comprender ella misma, no alcanza *de facto* a realizar un sistema de verdades absolutas y se ve obligada a modificar continuamente sus *verdades*, esto mismo demuestra que persigue la idea de la verdad absoluta, o de la auténtica verdad científica, y que vive en consecuencia dentro de un horizonte infinito de aproximaciones que aspiran a tocar esta idea. (Husserl 1963, 53).

Una vez en posesión de las ideas reguladoras de evidencia perfecta y de verdad absoluta, su uso por parte de Husserl para construir la ciencia absolutamente fundamentada, la filosofía--de la que tenemos hasta ahora tan sólo su idea, pero no su posibilidad efectiva--, se hace patente en la siguiente reflexión. Todas las verdades de la ciencia primera--y si no todas, sí al menos todas aquellas que hayan de formar parte esencial de su punto de partida--, han de ser verdades absolutas dadas con evidencia perfecta. Con esta formulación, el filósofo puede concretar ahora la meta a la que dirige todos sus esfuerzos de la siguiente manera: no admitiré como válido, "ningún juicio que no haya sacado de las fuentes de la evidencia" (Husserl 1963, 54). Como puede observarse, o al menos eso creo, una de las notas más características e importantes de esta etapa madura del pensamiento de Husserl a que da lugar la exposición previa, es que este filósofo ha transvasado e impuesto a las verdades que han de constituir el punto de

partida de la filosofía, lo que se descubre que es tan sólo la idea reguladora de las ciencias especiales. Es decir, Husserl ha puesto al principio de la ciencia filosófica como exigencia de su tarea lo que se descubre que está al final de las demás ciencias, al menos de las naturales. Este salto hacia atrás me parece fundamental. No obstante, parece ser que únicamente colocando esta exigencia de una evidencia perfecta y una verdad absoluta al comienzo de la ciencia filosófica, es posible que la determinación de la voluntad del filósofo se dé por satisfecha.

Ahora bien, ¿es realmente factible realizar lo que la voluntad se ha propuesto gracias a lo que el entendimiento acepta como posible, como algo que no encierra contradicción en sí? Es palpable que si no se consigue encontrar ningún conocimiento que cumpla todas las exigencias con las que Husserl ha cargado los conocimientos propios de la filosofía, entonces la realidad de la ciencia primera exclusivamente puede alimentar el deseo, pero nunca saciarlo. Sin embargo, antes de dar una satisfacción a la voluntad importa delimitar con cierta exactitud cuáles serían las propiedades de los conocimientos que formen el punto de partida de la filosofía. A juicio de Husserl estas propiedades son, como mínimo, las siguientes:

(i) El punto de partida debe proveer un conocimiento "primero en sí" que sirva de cimiento para todo el sistema de los conocimientos. La expresión "primero en sí" sólo puede querer decir aquí "absolutamente indubitable".

(ii) Este principio ha de ser "prácticamente realizable", es decir, ha de poderse encontrar y reconocer realmente este punto de partida, sea simple o múltiple.

(iii) La evidencia del punto de partida ha de ser anterior a todas las demás posibles evidencias.

(iv) Esta anterioridad no ha de ser sólo aparente, sino que ella misma se nos ha de dar también con evidencia. (Husserl 1963, 54-55).

Es en la discusión de estas propiedades del punto de partida de la filosofía donde tiene sentido pleno la idea de evidencia apodíctica (apodiktische Evidenz) que clasifica transversalmente tanto a la idea de evidencia adecuada como a la de evidencia inadecuada (Husserl 1963, 55). Como aquí interesa sobre todo la idea de evidencia apodíctica, será ésta la que analice, dejando de lado las otras dos especies posibles. Las dos notas esenciales de la evidencia apodíctica son, primero, que sea absolutamente imposible que lo que yo juzgo que existe no exista, o lo que yo juzgo que es *Y* no lo sea, y, segundo, que esta imposibilidad absoluta se me dé también de una forma evidente apodíctica (Husserl 1963, 56).

Después de contar con una idea algo más precisa de evidencia y, en concreto, de evidencia apodíctica, resta ante todo plantearse la pregunta de saber respecto de qué cosas podemos llegar a tener esta última clase de evidencia. Esta investigación, como se comprende de suyo, va de la mano con la necesidad de encontrar un principio o principios absolutos del conocimiento, así como de la noción de filosofía como ciencia rigurosa sin supuestos a la que el filósofo aspira. La primera aplicación de Husserl de la idea de evidencia apodíctica se refiere a la existencia del mundo. La segunda, como se anunció, a mi propia existencia y a mí mismo en tanto que sujeto que juzga.

Por lo que respecta a la primera aplicación de la idea de evidencia apodíctica, Husserl piensa que la existencia del mundo o, más exactamente, la proposición raramente formulada por evidente "el mundo existe", jamás podrá reclamar para sí la dignidad de cumplir cabalmente esta clase de evidencia (Husserl 1963, 57). Sus razones para pensar de esta manera ya las expuse en el capítulo tercero cuando reflexioné acerca del problema de la percepción, la existencia y el error. Estas razones son, como se recordará, básicamente cartesianas y apelan antes que nada a la posibilidad de principio, y no meramente de hecho, de que todo lo que se me da por

los sentidos sea siempre falso, ilusión, sueño. Un nuevo texto de Husserl servirá perfectamente bien para recordar estas tesis fundamentales y para constatar que siempre fue un acompañante fiel del idealismo transcendental fenomenológico—heredado acríticamente por Husserl de Descartes—la afirmación de que la tesis del mundo es siempre contingente.

Por lo que toca a lo primero <a su evidencia apodíctica> no puede reclamarse sin más para la experiencia sensible universal, en cuya evidencia nos está dado constantemente el mundo, una evidencia apodíctica, que en cuanto tal excluiría la posibilidad de tornarse dudosa la realidad del mundo o la posibilidad de la inexistencia de éste. No se trata sólo de que un objeto empírico determinado pueda padecer la desvaloración de convertirse en ilusión de los sentidos; también el conjunto entero de la experiencia abarcable como unidad en cada caso, puede revelarse y parece haberse revelado en Descartes como ilusión, como, por decirlo así, un sueño universal. (Husserl 1963, 57).

Sin embargo, tal y como traté de argumentar entonces, esto es justo lo que no puede pasar. Aludo aquí de nuevo a algo que se explicó detalladamente en aquella ocasión. Para recorrer el camino cartesiano de la duda hace falta haberse equivocado ya alguna vez o entender al menos el sentido de la experiencia del error. Pero una de las condiciones de posibilidad necesarias para dudar de la existencia de los objetos de los que tengo experiencias sensibles, es que alguna vez haya estado en contacto privilegiado con un trozo del mundo realmente existente. Por este motivo, únicamente es posible dudar de la existencia de objetos sensibles particulares, pero no de la existencia del mundo en su conjunto. Es así que cuando yo juzgo que el mundo existe, o al menos que ha existido, no puede suceder que no exista, aunque sí este o aquel objeto concreto. De esta forma, la proposición "el mundo existe" cumple al menos la primera nota de toda evidencia apodíctica: es absolutamente imposible que cuando yo juzgo que el mundo existe, nunca haya existido.

Por consiguiente, si bien creo que la argumentación de Husserl hasta el planteamiento de la evidencia apodíctica no tiene ninguna tacha, creo, por el contrario, que a partir de su crítica al grado de evidencia con la que se nos da la existencia del mundo en su conjunto la situación no es, ni con mucho, tan obvia. De todas las maneras, resulta palpable qué es lo que Husserl está dispuesto a extraer como consecuencia de su crítica a la evidencia empírica universal. Cuando se considera que la existencia del mundo no se nos da de una manera apodíctica, entonces hay que dejar en suspenso, entre paréntesis, tanto las evidencias del mundo de la vida con las que contamos a diario, como las evidencias de la ciencia natural. Si la existencia del mundo puede llegar a ser dudosa en virtud de la aplicación de la noción de evidencia apodíctica a la experiencia sensible global, entonces ninguna ciencia natural ni ninguno de sus datos empíricos, juicios o teorías puede admitirse como válidos, porque el suelo que les sirve como base permanente e indubitable, la existencia del mundo, no es de ahora en adelante un suelo absolutamente fiable, al menos desde el punto de vista de la exigencia de contar exclusivamente con evidencias apodícticas para la construcción de la filosofía primera. Por consiguiente, la actitud radical que ha querido adoptar el filósofo en el inicio de sus investigaciones lleva de suyo a "poner entre paréntesis" el mundo natural y todas las verdades referidas a él—especialmente todas aquellas que tienen a su base experiencias sensibles concretas—; en poner en tela de juicio, por tanto, "la creencia originaria" en el ser, en el valor de realidad del mundo y en su existencia.

En el caso del psicólogo convertido en el estilo de filósofo cartesiano, éste pondrá momentáneamente entre paréntesis, por difícil que le resulte en principio realizar esta "revolución cartesiana" (Husserl 1963, 48), su creencia en la validez de todos sus experimentos, de todas las teorías que ellos hayan podido originar fundadamente, de todos los supuestos heurísticos metodológicos que están involucrados en la elaboración

sobre las que se asienta el edificio entero de la ciencia. Una de estas presuposiciones sería, por ejemplo, contar con la evidencia de la existencia del mundo para elevar desde este cimiento el edificio entero de las ciencias. Esto es lo que hace la filosofía objetivista. Pero esto es precisamente lo que se prohíbe hacer Husserl en su búsqueda de la ciencia filosófica.

Ahora bien, si ni las evidencias más obvias del mundo de la vida, ni las más elejadas para el hombre común de la ciencia, son candidatos suficientemente robustos para formar parte de las proposiciones de la filosofía primera, de la que tampoco podría formar parte ninguna proposición que cualquier reflexión metateórica sobre las ciencias pudiera proveer, entonces ¿dónde encontrar los comienzos completamente indubitables del conocimiento que requiere Husserl? Sólo hay una "cosa" que la actitud radical del filósofo respecto de todos los conocimientos heredados jamás podrá poner entre paréntesis en su ser, en su realidad y en su existencia, porque es, precisamente, aquello que ha decidido adoptar libremente esa actitud e imponerse voluntariamente esa exigencia: yo mismo. Yo mismo y, especialmente, mi existencia y toda mi vida de conciencia, esto es, el fluir temporal de mis *cogitationes*, son elementos que quedan completamente asegurados en su valor y validez—apodícticamente, aunque quizás no adecuadamente, sobre todo por lo que se refiere a las vivencias pasadas—cuando decido realizar la epojé y la llevo realmente a cabo (Husserl 1963, 61). La vida filosófica consiste precisamente en darse cuenta del valor de fundamento que tiene el *ego cogito* para el conocimiento, para cualquier clase de conocimiento. También, por supuesto, para el conocimiento psicológico. Por consiguiente, es el yo del psicólogo que ha realizado la reducción transcendental el que se convierte así en el centro, en la fundamentación última de la psicología como ciencia objetiva de lo subjetivo, el único que puede entender la noción de "ser subjetividad" en su sentido pleno. El *ego* del psicólogo que ha recorrido el camino de la

de estas teorías, en la validez de los mismos datos empíricos sobre los que se apoyan éstas, hasta que no consiga fundamentarlos en conocimientos indubitables. A partir de ahora, nada de esto aceptará como válido si es que no consigue encontrar al menos un punto arquimédico cartesiano absolutamente indubitable a partir del cual justificar la validez de todos estos diversos elementos esenciales para su labor como científico natural.

Para ésta <la psicología científica> el mundo--el mundo presupuesto por ella como autoevidente y real--es el suelo; la epojé nos ha arrebatado precisamente este suelo. (Husserl 1976a, 182).

Por lo demás, esta estrategia necesaria de buscar conocimientos indubitables para la construcción de la filosofía como ciencia absolutamente fundamentada desde un inicio, no significa para Husserl que ese conocimiento o conocimientos indubitables haya que utilizarlos como premisas a partir de las cuales deducir con seguridad todos los demás conocimientos. Eso sería cartesianismo, contra el que Husserl reacciona enérgicamente (1963, 63).

Antes de seguir acompañando a Husserl en su búsqueda de la conciencia absoluta no quiero dejar de pasar por alto el lugar que él asigna a estos conocimientos indubitables en su última obra con el fin de redondear algunos puntos. Con terminología posterior de Husserl, estos conocimientos absolutamente indubitables son los que formarían parte principal de aquellas evidencias protooriginarias todavía anteriores a las evidencias del mundo de la vida y, por supuesto, a las evidencias de la ciencia de las que se habló anteriormente. El paso definitivo más característico de Husserl en *La crisis de la ciencias europeas y la fenomenología transcendental* y que se sigue de suyo del camino de la reducción transcendental descubierto años antes, es convertir en *presuposiciones* (*Voraussetzungen*) las propias evidencias del mundo de la vida

reducción trascendental para hacerse plenamente consciente de que él es también ego trascendental y no solamente ego empírico, psicofísico, es la conciencia absoluta que se buscaba, el "yo trascendental", y, en concreto, la posición absolutamente apodíctica que juzga la validez de toda su labor como científico natural, así como la de sus métodos, datos, juicios y resultados teóricos (Husserl 1963, 64-65). El *ego cogito* es, por añadido, "la base apodícticamente cierta y última de todo juicio, en que hay que fundamentar toda filosofía radical" (Husserl 1963, 58).

Con todo este camino ya andado, es posible reforzar algunas tesis anteriores que no estaban del todo justificadas, aunque se podían entender intuitivamente de sobra en su contexto. Al contraponer la conciencia empírica, objeto de la ciencia psicológica, con la conciencia absoluta, patrimonio exclusivo de la filosofía fenomenológica trascendental, comentaba, preveyendo sin duda lo que había de venir, que la conciencia absoluta, el *ego* trascendental, *no era una parte del mundo*. Pero es solamente ahora cuando se puede vislumbrar, si no entender con absoluta claridad, como sería deseable para Husserl, que semejante tesis se sigue de suyo de las argumentaciones previas. Si *todo* el mundo natural, sobre el que se vuelcan teórica y empíricamente las ciencias especiales, ha quedado puesto entre paréntesis en su validez de ser, entonces no queda *nada* perteneciente a él—ni este mundo natural completo, ni ninguna de sus partes—, con cuya validez, ser y existencia se pueda contar sin más ni más. Sin embargo, yo mismo y toda mi vida de conciencia permanezco inalterable en mi valor de ser y realidad. Por consiguiente, yo, al convertirme a mí mismo, después de realizar la *epojé*, en *ego* trascendental, no soy, ni puedo ser, parte del mundo natural. Este, creo, es el argumento decisivo para sortear aquella tesis.

Por otra parte, para el fundador de la fenomenología es patente que la investigación sistemática de todas estas evidencias protooriginarias

anteriores y presupuestas por las demás evidencias--las del mundo de la vida y las del mundo de la ciencia--, forman también de por sí un ámbito de investigación científica cerrado que, obviamente, está animado por una noción de cientificidad diferente de la empleada en la ciencia natural, pero no por ello menos legítima. De hecho, para Husserl es la noción de ciencia más importante y originaria de todas, porque es la única que se pone a sí misma la exigencia de contar con evidencias apodícticas--verdadero fin de la razón que marca el rumbo de las ciencias especiales--, en el comienzo de sus investigaciones. Todas las demás formas de hacer ciencia son siempre secundarias respecto de ésta. Además, estas dos nociones posibles de ciencia respaldan a su vez dos formas radicalmente diferentes de entender la filosofía desde la modernidad y que han sido su motor constante.

Toda la filosofía moderna en su sentido originario como ciencia universal, absolutamente fundamentada es, de acuerdo con nuestra descripción, por lo menos desde Kant y Hume, una única lucha entre dos ideas de ciencia: la idea de una ciencia objetivista sobre el suelo del mundo previamente dado, y aquella otra idea de una filosofía sobre el suelo de la subjetividad absoluta transcendental. (Husserl 1976a, 212).

& 8. La crisis de la fenomenología transcendental y las ciencias europeas.

A pesar de todas estas aclaraciones quedan todavía importantes problemas por delante nada fáciles de resolver. El psicólogo, aún después de realizar la reducción transcendental y de encontrar posteriormente que él "se pone a sí mismo exclusivamente como *fundamento del valor* de todos los fundamentos y valores objetivos" (Husserl 1963, 64-65), se encuentra de todas las maneras, como datos precisamente encerrados en su vida de

conciencia, con la tarea posible, y más aún, necesaria, de desarrollar una ciencia, su ciencia, con los datos empíricos que se desprenden de ciertos experimentos cuidadosamente diseñados, con ciertos métodos interpretativos, con juicios y teorías y, por lo tanto, también con el mundo en un sentido muy peculiar, con el mundo de las teorías. Aunque mi *ego*, más allá de mi labor como psicólogo, haya puesto el ser y la existencia del mundo, gracias al cual recojo todos mis datos empíricos, entre paréntesis, sin embargo "el mundo experimentado en esta vida reflexiva sigue siendo para mí "experimentado" en cierto modo, y *exactamente con el mismo contenido peculiar que antes*. Continúa siendo el fenómeno que era antes, sólo que yo, en cuanto que reflexiono filosóficamente, ya no llevo a cabo, ya no concedo validez a la creencia natural en la realidad que es inherente a la experiencia del mundo" (Husserl 1963, 59; el subrayado es mío).

¿Qué significa esta afirmación tan rara de que el mundo sigue teniendo el mismo contenido para mí, aunque, por haber hecho yo la epojé en tanto que sujeto transcendental, no le conceda ya valor de realidad? ¿Qué consecuencias tiene? En la tercera crítica que hace Iso Kern al camino cartesiano de la reducción transcendental, el que se ha seguido especialmente hasta aquí, este autor argumenta--quizás de una forma algo exagerada, aunque aquí no voy a discutir este punto--que sólo la forma temporal de mis vivencias, es decir, su organización incesante en la estructura formal inamovible de presente, pasado y futuro, y mi propia existencia, son los únicos elementos apodícticos que puedo obtener de la epojé que transcurre por el camino cartesiano. Sin embargo, estos dos elementos no tienen en realidad ningún contenido concreto a partir del cual elaborar una ciencia, ni siquiera la filosófica (Kern 1962, 321). Contenidos, al parecer, sólo los puede proveer el mundo, aunque éste sea considerado ahora por mí, en tanto que *ego* transcendental, como un mero "fenómeno" sobre el que juzgo, y no como algo real y existente con pleno derecho

desde un inicio. El mundo, de esta manera, sigue manteniendo exactamente el mismo contenido que siempre. Nada en este sentido se ha alterado. Pero al practicar la reducción trascendental yo lo transformo en el "fenómeno" que se me da constantemente en mis *cogitationes y sobre el que yo juzgo, valoro, etc.*

En la reorientación (*Umstellung*) de la epojé nada se pierde, nada de todos los intereses y fines de la vida en el mundo y, de igual modo, tampoco se pierde nada de los fines cognoscitivos, sólo que para todo ello se muestra sus esenciales correlatos subjetivos, por lo cual se expone el pleno y verdadero sentido de ser del ser objetivo y, de este modo, de toda verdad objetiva. (Husserl 1976a, 185).

A la luz de todos estos elementos teóricos singulares hay que preguntar de nuevo qué es lo que, en verdad, ha ganado el psicólogo en relación a todos los problemas de fundamento de la ciencia psicológica; qué ha ganado, por ejemplo, en relación a la justificación de la validez de los métodos de la psicología; qué en relación a la discusión de las relaciones entre estados psicológicos y estados físicos; qué en relación a los modelos computacionales que adopta aquella metodología en alguna de sus fases; qué en relación a la validez de las teorías que se alcanza finalmente a construir; qué en relación a los compromisos ontológicos de estas teorías. Por lo que respecta a la cuestión que se refiere a la validez de las teorías científicas naturales siempre parecerá más plausible pensar que lo que las justifica es la investigación empírica progresiva, que la búsqueda de un principio absolutamente apodíctico del conocimiento, búsqueda que encontraría su descanso en la completa indubitabilidad que tiene para mí mi propia existencia y toda mi vida de conciencia. No es de extrañar, en este sentido, que Husserl se preguntara si el espíritu de la filosofía cartesiana que él quería reivindicar, corregir, ampliar y potenciar al máximo, era apropiado "para infundir fuerzas vitales a nuestro tiempo" (Husserl 1963,

45-46). En especial, lo que más le llamaba la atención, desde su punto de vista filosófico, era el hecho de que cada vez que las ciencias especiales conseguían preocuparse por las cuestiones de fundamento que las atosigaban nunca, o en muy pocas ocasiones, se les ocurriera volver la mirada a reflexiones teñidas de espíritu cartesiano. Lo curioso, según Husserl, es que las ciencias, al proceder así, se alejarían cada vez más y más de la única clase de investigaciones que conseguiría fundamentarlas de una vez y para siempre.

Sin embargo, no alcanzo a ver del todo que la situación haya que describirla exactamente como lo hace Husserl. En concreto, no alcanzo a ver con suficiente claridad que el punto de partida absolutamente indubitable de *mi* conocimiento, sea, a su vez, el punto de partida de las ciencias especiales. Si hago esta afirmación es porque hay que darse cuenta de que este o estos puntos de partida indubitables no pueden ser el fundamento de *todos* los conocimientos. Al menos los conocimientos de la naturaleza y de todo lo que a ella pertenece como, por ejemplo, yo mismo en tanto que sujeto psicofísico, tarde o temprano están expuestos a la marcha incesante de las conjeturas y las refutaciones. Por eso hablo del punto de partida de mi conocimiento, pero no del conocimiento. Esto tiene también que ver con el hecho de que me parece enteramente necesario distinguir entre el fundamento inteligible-mundo de la vida-y el fundamento ontológico de una ciencia, este último quizás sólo alcanzable en el progreso infinito de la investigación. En cualquier caso, ninguna investigación encerrada en el ego transcendental husserliano podría jamás dar cuenta cabal de esta diferencia sin salir de la investigación fenomenológica o, al menos, sin forzarla excesivamente en sus pretensiones originales.

No obstante, ésta no es la única desventaja que tiene la fenomenología transcendental respecto de las ciencias especiales y de su manera de proceder objetivista. Otro ejemplo ilustrará las nuevas

dificultades. Según Husserl, lo más característico del *ego* del filósofo, o del psicólogo o de cada cual, es que no podrá ser identificado nunca enteramente con el hombre natural, con el hombre psicofísico, como se ha insistido en repetidas ocasiones. El ego transcendental convierte al mundo y a todo lo que hay en él, *incluido él mismo como objeto de la psicología*, en un puro "fenómeno" sobre el que juzga. En este sentido, no es de extrañar que el psicólogo de Dennett, el heterofenomenólogo, convierta al yo del sujeto de experimentación en algo que se le aparece, en un fenómeno, en un texto psicológico que hay que interpretar de acuerdo con ciertos procedimientos heurísticos que se aceptan como racionalmente válidos. Como decía también Husserl "los demás hombres y los animales sólo son para mí datos de la experiencia (Erfahrungsgegebenheiten)" (Husserl 1963, 58). De cualquier forma, no hay que confundir el hecho de que el psicólogo convierta al yo del otro en un fenómeno investigable científicamente, con el hecho, todavía más radical y difícil de entender, de que el yo transcendental, al colocar entre paréntesis las evidencias referidas al mundo y, por lo tanto, al considerar el mundo sólo como correlato, se convierta a sí mismo, como sujeto empírico, como sujeto en el mundo, y a su conciencia, como conciencia igualmente empírica, en correlatos, en "fenómenos" de su yo transcendental (Husserl 1963, 65). Pero si es verdad esta tesis, es decir, si la conciencia empírica es siempre "correlato" de la conciencia absoluta, entonces la teoría fenomenológica de las relaciones entre la conciencia empírica y la conciencia transcendental se complica—por no decir que se oscurece—hasta límites casi indefinidos. Bastaría simplemente llamar aquí la atención sobre la tesis de Husserl que sostiene que, en cierto sentido, no hay ninguna diferencia entre el sujeto empírico y el sujeto transcendental, para hacer palpables estas dificultades supremas.

Con todo, psicología y filosofía transcendental se encuentran hermanadas entre sí de forma peculiar e inseparable, a saber: en

virtud del hermanamiento... de diferencia e identidad entre el yo psicológico (así, pues, humano, mundanizado en el mundo espacio-temporal) y el yo transcendental. (1976a, 209).

Cuando se reúnen todos estos diversos elementos no se puede dejar de caer en la cuenta de que en este ámbito hay algo especialmente difícil que Husserl nunca terminó de dominar. Hay un abismo que separa irremediablemente al yo transcendental del yo empírico. En primer lugar, aquél no es nunca parte del mundo, mientras que éste sí que lo es. Por este motivo, el yo empírico siempre es investigable con los métodos de la ciencia natural. El yo transcendental, por el contrario y por principio, nunca lo sería. Este yo transcendental es también el fundamento de valor del yo empírico. El yo empírico, sin embargo, es simplemente el "correlato" del yo transcendental. Todas estas diferencias son de radical importancia en la filosofía primera de Husserl. Son, además, las únicas que permitirían ir más allá de la investigación del ser psíquico que propone la filosofía objetivista. Sin embargo, en algún otro sentido, "ambos yoes" son el mismo yo. Ambos soy yo.

¿Por qué dijo Husserl esto? ¿Qué necesidad tenía de decirlo? Su necesidad admite una hipótesis explicativa. Si no se impusiera la identidad del yo empírico con el yo transcendental, a pesar de todas sus diferencias, entonces sería inexplicable cómo podría darse la relación entre la conciencia y el cuerpo vivo sobre el que ella impera y en el que se hace cuerpo. En el fondo, es el cuerpo propio en el mundo natural lo que constituye el dato más indómito que existe para la fenomenología. Por el contrario, es precisamente este cuerpo, al menos en cierto sentido, el que favorecería las investigaciones de la psicología empírica. No hay psicología empírica de yoes sin cuerpo.

Redondear y finalizar el curso de la argumentación que pretende hacer resaltar las difíciles relaciones que se dan entre la fenomenología

transcendental y la psicología empírica, es algo que se puede hacer perfectamente aludiendo a este texto central de las *Meditaciones Cartesianas*:

...cuando digo que yo existo, *ego cogito*, esto ya no significa que existo yo, este ser humano. Yo no soy ya... el ser humano que se encuentra con su propia pura *mens sive intellectus sive animus*... Apercebidos de este "modo natural", somos yo y todos los demás hombres temas de las ciencias positivas u objetivas en el sentido corriente... también de la psicología. La vida psíquica de que habla la psicología ha sido entendida en todo tiempo, y sigue siéndolo, como vida psíquica en el mundo... Pero la *epojé* (ἐποχή) fenomenológica, que requiere del sujeto que filosofa la marcha de las meditaciones cartesianas depuradas, inhibe el valor de realidad del mundo objetivo... Para mí, el yo que medita, el yo que encontrándose y permaneciendo en la *epojé* (ἐποχή) se pone a sí mismo como *fundamento de valor* de todos los fundamentos y valores objetivos, no hay, pues, ni yo psicológico, ni fenómenos psíquicos en el sentido de la psicología, esto es, como partes integrantes de seres psicofísicos humanos. (Husserl 1963, 64-65).

A lo largo de todas estas meditaciones, que han tenido como objeto de estudio la filosofía de la psicología de Husserl, y sobre todo teniendo un texto como éste ante la vista, se hace cada vez más claro que Husserl ha considerado al sujeto transcendental como *una existencia que juzga otras existencias y sus propiedades reales*. Es especialmente esta existencia indubitable del sujeto transcendental, en lo que tiene de apodíctica, y su propiedad de juzgar y valorar el mundo, lo que haría de ese yo transcendental, de mí mismo, algo no investigable científicamente. Sin embargo, Husserl considera que el sujeto transcendental que es cada cual, la existencia que juzga y valora otras existencias, está en una conexión sin solución de continuidad con la aprehensión originaria que hace de sí misma y de los demás como existencias psicofísicas en el mundo espacio-temporal y causal. Hay un hermanamiento de identidad entre ambas existencias, como comentaba

Husserl. Pero yo, como existencia psicofísica, puedo ser el objeto de investigación de un psicólogo en la forma reseñada una y otra vez.

& 9. La conciencia absoluta en el mundo.

Como se ha visto, una forma provisional de dar sentido a la idea de que el sujeto no es parte del mundo es decir de él que es "la posición absolutamente apodíctica" desde la que se juzga el mundo. La posición absoluta que, por ejemplo, juzga acerca de la identidad o no de los sucesos psicológicos y físicos. Este sujeto siempre podría interpretarse a sí mismo como una parte del mundo. No que él sea siempre una parte del mundo y que no quepa nunca salir del círculo de problemas que plantea el hecho de que él siempre es una parte del mundo, sino que el sujeto absoluto se interprete, quizá necesariamente, como una parte del mundo, es lo realmente problemático y decisivo.

En relación con ello, a este rendimiento (Leistung) constituyente del mundo pertenece esencialmente el hecho de que la subjetividad se objetiva a sí misma como subjetividad humana, como parte integrante del mundo (Die Subjektivität sich selbst als menschliche, als Bestand del Welt, objektiviert). (Husserl 1976a, 115-116).

En cierto sentido, hay que decir que para Husserl esta interpretación, si bien necesaria, no es primera en la consideración del sujeto. Sólo una vez que el sujeto se ha interpretado a sí mismo como parte del mundo cabe que se haga una ciencia "objetiva" de lo subjetivo. La psicología es, por consiguiente, en el orden de la fundamentación, un producto derivado en el sentido de que, previamente a ella, en un sentido metodológico fundamental, siempre hay que considerar al sujeto absoluto que es quien lleva a cabo últimamente esa interpretación como un movimiento necesario suyo. Lo que está diciendo Husserl es que el psicólogo debe ser ese sujeto

absoluto a partir del cual cobra sentido claro que él está haciendo ciencia objetiva de lo subjetivo. El psicólogo no podría entender qué es subjetividad, acción y pasión del yo, si no se retrotrae a sí mismo como posición absoluta apodíctica desde la que juzgar todo lo referente a la subjetividad, a cualquier forma de subjetividad y a su posible inserción necesaria en el mundo, a su "objetividad", a su carácter de posible objeto de ciencia. Y esa posición no es una parte del mundo. Esta es, insisto una vez más, la tesis última de Husserl, la más fundamental de todas. Otra forma de decir esto mismo es que el sujeto transcendental no es, ni puede ser, tema de ninguna ciencia natural. Defender que sí que lo es sería un puro contrasentido².

Creo, sin embargo, que esta tesis de Husserl, sin ser claramente falsa--¿quién negaría que cuando uno se dice subjetividad se dice a sí mismo de una forma única y apodíctica?--encierra la dificultad suprema de explicar el "tránsito" necesario del sujeto transcendental a parte efectiva del mundo objetivo. Por mucho que su inserción en el mundo como parte objetiva suya sea una interpretación--si bien necesaria--que lleva a cabo la subjetividad transcendental respecto de sí misma, esa interpretación carecería de toda "verosimilitud"--la verosimilitud es lo característico de toda interpretación--si no hay una base objetiva suficiente para formularla. La interpretación de una poesía puede oscilar entre diversos grados de verosimilitud que se van mejorando con las diversas lecturas a lo largo del tiempo individual e histórico. No hay *una* interpretación, pero hay interpretaciones absolutamente insostenibles. En este caso, sucedería algo semejante. Si el sujeto transcendental se interpreta y objetiva como parte del mundo, como parte del orden causal natural, es porque esa interpretación es, cuanto menos, verosímil. Pero es tan verosímil que tiene visos de ser casi una

² Thomas Nagel (1970) y Collin McGinn (1991) han defendido una doctrina muy semejante, aunque ni con mucho tan radical y tan bien planteada como la de Husserl.

autoevidencia. Ella parece ser al menos una de esas autoevidencias del mundo de la vida. Y lo que hace falta explicar es precisamente cómo es posible esa inserción, esa interpretación. Lo que se propone en filosofía de la mente contemporánea, es que la explicación de esa inserción y la clave de esa interpretación sólo se pueden obtener adoptando la posición metateórica respecto de la mejor ciencia psicológica con la que contamos. Y esa psicología, por lo que se refiere exclusivamente a los procesos cognitivos, se llama psicología cognitiva.

Para afianzar este punto, que en el fondo no es más que otro aspecto del problema de las cinestesis, sostendré que numerosas expresiones de las que se sirve Husserl para aludir a la manera en que se podría pensar las relaciones entre la conciencia y el cuerpo vivo resultan extremadamente insuficientes. Las expresiones "interpretación" del sujeto transcendental como parte del mundo, "objetivación" de la subjetividad como subjetividad humana, subjetividad empírica humana como "correlato" de la subjetividad transcendental, "participación" (*Teilnahme*) de la subjetividad en la corporalidad, corporalización (*Verkörperung*), son tan sólo títulos de problemas de la filosofía fenomenológica que necesitan completarse con una teoría psicológica natural y objetiva adecuada. En ningún caso son soluciones, sino rótulos para posibles ámbitos de investigación. No es de extrañar que Husserl adjetivara de *paradójico* el problema de las relaciones entre la conciencia y el mundo material con su orden espacial, temporal y causal.

La paradoja de la subjetividad humana: el ser-sujeto para el mundo y, al mismo tiempo, ser-objeto en el mundo. (Husserl 1976a, 182).

También, por ejemplo, en el texto del ejemplar III de *Ideas* reconoce que la "subjetividad psicofísica" plantea un orden de dificultades especiales (*Ideas* 474). Para que estas afirmaciones, que pretenden criticar ciertos

aspectos de la teoría de Husserl, no queden damasiado en el aire, es importante formular de nuevo la pregunta "¿cómo la conciencia puede hacerse mundo, ser parte del mundo?", y ofrecer a continuación la clase de respuesta que puede encontrarse en sus textos principales para que se vea su insuficiencia.

Advertimos enseguida que sólo lo puede mediante una cierta *participación* (*Teilnahme*) en la trascendencia en su primer y originario sentido, y éste es patentemente el de la trascendencia de la naturaleza material. Sólo por su relación empírica con el cuerpo vivo (*Leibe*) se convierte la conciencia en realmente humana o animal, y sólo por ese medio ocupa un lugar en el espacio y en el tiempo de la naturaleza—en el tiempo que se mide físicamente. Acordémonos también de que sólo mediante el enlace de la conciencia y el cuerpo vivo (*Verknüpfung von Bewusstsein und körperlichem Leib*) y con una unidad intuitivo-empírica natural, es posible algo así como una comprensión mutua entre los seres animados pertenecientes a un mundo, y que sólo por este medio puede cada sujeto cognoscente encontrarse con el mundo en su plenitud, consigo mismo y con otros sujetos, y a la vez reconocer en él un mismo mundo circundante, común a él y a todos los demás sujetos (Husserl 1950, 130-131) (El subrayado es mío).

Pero el problema esencialísimo con el que se enfrenta la filosofía de la mente es la tensión creada entre decirse sí mismo y subjetividad de una forma única y apodíctica y la teoría, siempre provisional, que ha de llenar de una manera comprensible las "metáforas" de la participación, de la interpretación y de la objetivación de la subjetividad. En este sentido se le asigna al alma "un tipo de ser principalmente análogo al de la naturaleza" (Husserl 1976a, 64).

& 10. El dualismo epistemológico y el dualismo de las sustancias.

Todo apunta a que hemos llegado a un punto especialmente conflictivo y difícil. Estas últimas reflexiones indican de nuevo la necesidad de una ciencia psicológica natural "objetivista" que extraería su sentido como ciencia de la reflexión acerca del ser psíquico en el mundo de la vida que le serviría como su fundamento inteligible. Paralelamente a esta investigación natural, y cuando la ciencia psicológica se encuentre en cierto estadio de madurez, existiría constantemente la posibilidad de una reflexión metateórica acerca de los métodos, teorías, resultados, compromisos ontológicos y consecuencias de la ciencia psicológica así construida paso a paso. Pero esta obligatoriedad y esta marcha de los esfuerzos teóricos desde la psicología popular a la psicología científica, tienen para Husserl un freno inevitable que tiñe todas las investigaciones en torno a la subjetividad con un matiz nuevo. Este "obstáculo" es mi yo absoluto, transcendental, desmundanizado, que es el que juzga siempre acerca de estos métodos, de estas teorías, de estos resultados, etc., y que no puede ser objeto de ninguna ciencia natural.

Por la manera en que es obligatorio proceder para aclarar qué es este yo transcendental, es posible destacar dos de las funciones esenciales que él cumpliría: una función de fundamentación epistemológica y una función de fundamentación ontológica. Es así que entre (i) las evidencias protooriginarias, (ii) el mundo de la vida y sus evidencias y (iii) el mundo de la ciencia y las suyas, se darían curiosas relaciones quizás no del todo vistas por Husserl. Respecto del grado de evidencia de unas y otras evidencias, las evidencias protooriginarias tienen la mayor dignidad y certeza pensable. Y entre esta clase de evidencia últimas, especialmente mi existencia y su capacidad de juzgar y valorar son dos especialmente claras. Habría aquí una especie de *dualismo epistemológico*, o, al menos, de ruptura radical

entre estas evidencias y el grado de evidencia propio de los enunciados del mundo de la vida y de la ciencia. En este sentido, es importantísimo recordar aquí que Rorty considera que el último refugio antipragmatista que cabría seguir defendiendo en contra del empuje de la ciencias cognitivas, es la defensa de la tesis cartesiana de que la subjetividad es el único lugar seguro en el que poder hallar conocimientos seguros e indubitables (Rorty 1982b, 187).

Este reducto, no obstante, no puede mantenerse aislado, porque en seguida surge el problema de entender cómo esa cara del yo existente que juzga el mundo tiene otra cara que sí es mundo, sujeto empírico. En la investigación de esta otra cara no habría tanto dualismo epistemológico, sino, ante todo, un desconocimiento enorme. El deseo de conocimiento en este terreno es el que quiere satisfacer la psicología y las ciencias cognitivas, y el que quiere poner en claro el pensador de orientación metateórica. Por su parte, la investigación metateórica sobre los compromisos ontológicos de la psicología cognitiva, de creer a Fodor, sólo nos comprometerían con el dualismo de las propiedades físicas y psicológicas, pero no con el dualismo de las sustancias. Ontológicamente estaríamos comprometidos con el materialismo.

Es decir, habría en el fondo un desconocimiento absoluto respecto del "alma", respecto de mí mismo como sujeto consciente en el mundo, cuando me atengo meramente a la indubitabilidad de los conocimientos indubitables. Pensando en esta tensión es por lo que dije antes que ni los fenomenólogos tenían tantos motivos para estar contentos y desdeñar la investigación cognitiva actual, ni estos últimos tenían por qué sentirse tan intranquilos cuando se les habla de fenomenología.

En este sentido, la evidencia "yo existo" que se ha alcanzado mediante el método de la duda radical puede que sea una pieza clave para la teoría del conocimiento, cuyas posibilidades más fecundas sacó a la luz

la fenomenología de Husserl. Sin embargo, esta pieza es inerte para la filosofía de la mente y para la ontología de lo mental. De ser así, entonces no habría ninguna continuidad de fundamentación ontológica entre las evidencias del mundo de la vida--a las que seguirían las evidencias de la ciencia--y las evidencias protooriginarias a las que aspira la filosofía primera concebida a la manera husserliana, sino que, por el contrario, habría una ruptura insuperable. En este sentido, no se estaría estudiando jamás lo mismo desde la fenomenología transcendental cuando ésta dice yo, es decir, cuando yo digo yo absoluto después de haber recorrido el camino de la reducción transcendental, que desde la psicología científica cuando ésta dice "aquí hay un yo". Lo que no puede suceder, por consiguiente, es lo que dice Husserl en estas palabras:

Yo, en tanto que hombre y alma humana me convierto en primer lugar en tema de la psicofísica y de la psicología; pero a continuación en una nueva y más elevada dimensión, en tema transcendental (1976a, 209).

Son motivos como éstos los que podrían haber llevado a Rorty a sostener que "no tenemos ningún dato acerca de la naturaleza de lo mental. En particular, no tenemos ningún conocimiento de lo que es tener una mente cuando miramos dentro y hacemos introspección. Esta negación de que la mente (o "conciencia" o "subjetividad") es un género natural... es característica de una tradición que ha sido dominante en la filosofía anglosajona durante las tres últimas décadas" (Rorty 1982, 325).

¿No será importante distinguir aquí la ontología del sujeto psicofísico, el único sujeto del que se puede hacer realmente una ontología "científica" y, en verdad, quizás la única ontología posible, y ciertas cuestiones epistemológicas relacionadas con la certeza de su existencia, en

el sentido de Rorty? ¿Cómo podría haber una ontología de algo que no tiene ningún contenido?

En ningún momento de la historia de la psicología ha habido una teoría psicológica sobre la que se pudieran contrastar mejor las tesis de la fenomenología de Husserl a propósito de las relaciones entre fenomenología transcendental y psicología--entre Ego y alma--, como con la psicología cognitiva. Este hecho dice mucho tanto en favor de la psicología cognitiva, como de la fenomenología. Nunca habían estado las cosas tan claras como hasta ahora. Pero la psicología objetiva sigue siendo el verdadero campo de las decisiones en lo que a la ontología ser psíquico respecta (Husserl 1976a, 212).

* * * *

¿Que cuál es la tesis de este libro? Quizás la única posible, aquella que resumen las palabras que Galdós solía repetir al final de sus días: "¡Cuánto misterio!".

BIBLIOGRAFIA

- Acero, Juan José. 1990. "Platón, el lenguaje y el marciano." Revista de Filosofía. Madrid: Editorial Complutense. 3ª Epoca. Vol III, 3: 205-210.
- . 1991. "Después del análisis: significado, comprensión e intencionalidad." En Filosofía analítica hoy. Encuentro de tradiciones. Mercedes Torrevejano (Coord.). Santiago de Compostela: Universidad de Santiago de Compostela: 9-26.
- Achinstein, Peter. 1979. "The Causal Relation." Midwest Studies in Philosophy. Studies in Metaphysics. Ed. Peter A. French et al. Minneapolis: University of Minnesota Press: 369-386.
- Anderson, Alan Ross, ed. 1964. Mind and Machines. Englewood Cliffs: Prentice Hall.
- Anderson, David. 1987. "Is the Chinese Room the Real Thing?" Philosophy 62: 389-393.
- Anscombe, G.E.M. 1981. "The Intentionality of Sensation. A Grammatical Feature." Metaphysics and the Philosophy of Mind: The Collected Philosophical Papers. Volume II. Oxford: Basil Blackwell: 3-20.
- Aquila, Richard E. 1981. "'De Re, de Dicto', and Naturalism." The Journal of Philosophy: 718-719.
- . 1982. "On Intensionalizing Husserl's Intentions." Nous 16: 209-226.
- . 1989. "Intentionality, Content and Primitive Mental Directedness." Philosophy and Phenomenological Research 49: 583-604.
- Aristóteles. De Anima. Traducción e introducción de Tomás Calvo Martínez. Madrid: Biblioteca Clásica Gredos.
- Armstrong, D.M. 1968. A Materialist Theory of the Mind. London: Routledge & Kegan Paul.
- . 1977. "The Causal Theory of the Mind." Neue Hefte für Philosophie. 11: 82-95.
- . 1983. "Recent Work on the Relation of Mind and Brain." Contemporary Philosophy: A New Survey. Ed. Guttorm Floistad. The Hague: Martinus Nijhoff: 45-79.
- Bacon, John. 1986. "Supervenience, Necessary Coextension, and Reducibility." Philosophical Studies 49: 163-176.

- Barwise Jon y John Perry. 1983. Situations and Attitudes. Cambridge, Mass.: The MIT Press.
- Bechtel, William. 1988. Philosophy of Mind: An Overview for Cognitive Science. Hillsdale, NJ.: Lawrence Erlbaum Associates.
- Beckermann, Ansgar. 1988. "Language Understanding Machines." Erkenntnis 28: 65-85.
- Bennett, Jonathan. 1982. "Psychology and Semantics: Comments on Schiffer's "Intention-Based Semantics"." Notre Dame Journal of Formal Logic 23: 258-262.
- Bilgrami, Akeel. 1987. "An Externalist Account of Psychological Content." Philosophical Topics 15: 191-226.
- . 1989. "Realism Without Internalism: A Critique of Searle on Intentionality." The Journal of Philosophy 86: 57-72.
- Block, Ned. J. y Jerry A. Fodor. 1972. "What Psychological States Are Not." Philosophical Review April 1972: 159-181.
- Bogdan, Radu J. 1989. "Does Semantic Run the Psyche?" Philosophy and Phenomenological Research 49: 687-700.
- Boghossian, Paul A. 1990. "The Status of Content." The Philosophical Review 99: 157-184.
- Bradie, Michael P. 1976. "The Causal Theory of Perception." Synthese 33: 41-47.
- Brentano, Franz. 1973. Psychologie vom empirischen Standpunkt. 2 vols. Ed. Oskar Kraus. Hamburg: Felix Meiner.
- . 1982. Deskriptive Psychologie. Hamburg: Felix Meiner.
- Burge, Tyler. 1979. "Individualism and the Mental." Midwest Studies in Philosophy. Studies in Metaphysics. Ed. Peter A. French et al. Minneapolis: University of Minnesota Press: 73-121.
- . 1982. "Two Thought Experiments Reviewed." Notre Dame Journal of Formal Logic 23: 284-293.
- . 1989. "Individualism and Causation in Psychology." Pacific Philosophical Quarterly 70: 303-322.
- Bynum, Terrell Ward. 1985. "Artificial Intelligence, Biology, and Intentional States." Metaphilosophy 16: 355-377.
- Carleton, Lawrence R. 1984. "Programs, Language Understanding, and Searle." Synthese 59: 219-230.
- Carnap, Rudolf. 1965. "Psicología en lenguaje fisicalista". El positivismo lógico. Ed. A. J. Ayer. México: Fondo de Cultura Económica: 171-204.
- . 1972. "Logical Foundations of the Unity of Science." Intentionality, Mind and Language. Ed. Ausonio Marras. Urbana, Chicago, London: University of Illinois Press.
- Cartwright, Nancy. 1979. "Do Token-Token Identity Theories Show Why We Don't Need Reductionism?" Philosophical Studies 36: 85-90.

- Castañeda, Hector-Neri. 1977. "Perception, Belief, and the Structure of Physical Objects and Consciousness." Synthese 35: 285-351.
- Clark, Romane. 1973. "Sensous Judgments." Nous 7: 45-46.
- Cole, David. 1984. "Thought and Thought Experiment." Philosophical Studies 45: 431-444.
- Collins, Arthur W. 1979. "Could Our Beliefs Be Representations in Our Brains?" The Journal of Philosophy 76: 225-243.
- Coppock, Paul. 1987. "Putnam's Transcendental Argument." Pacific Philosophical Quarterly 68: 14-28.
- Craig, Edward. 1976. "Sensory Experience and the Foundations of Knowledge." Synthese 33: 1-24.
- Cuda, Tom. 1985. "Against Neural Chauvinism." Philosophical Studies 48: 111-127.
- Cummins, Robert. 1975. "Functional Analysis." The Journal of Philosophy 72: 741-765.
- . 1981. "What Can Be Learned from *Brainstorms*?" Philosophical Topics 12: 83-92.
- . 1984. The Nature of Psychological Explanation. Cambridge, Mass.: The MIT Press.
- . 1989. Meaning and Mental Representation. Cambridge, Mass.: The MIT Press.
- Chacón, Pedro. 1992. "Por qué pude escribir este artículo. Reflexiones sobre el fantasma de lo mental." Revista de Filosofía 3ª época 5: 119-141.
- Chihara, C. S. y J. A. Fodor. 1991. "Operationalism and Ordinary Language: A Critique of Wittgenstein." The Nature of Mind. Ed. David M. Rosenthal. Oxford: Oxford University Press: 137-150.
- Chisholm, Roderick M. 1958. "Intentionality and the Mental." Minnesota Studies in the Philosophy of Science. v.2. Ed. Herbert Feigl, Michael Scriven, y Grover Maxwell.
- . 1967. "Brentano on Descriptive Psychology and the Intentional." Phenomenology and the Existentialism. Ed. N. Lee y Maurice Mandelbaum. Baltimore: The Johns Hopkins Press: 1-23.
- . 1972. "Sentences about Believing." Intentionality, Mind and Language. Ed. Ausonio Marras. Urbana, Chicago, London: University of Illinois Press.
- . 1981. "Brentano's Analysis of the Consciousness of Time." Midwest Studies in Philosophy, The Foundations of Analitic Philosophy. Ed. Peter A. French et al. Minneapolis: University of Minnesota Press: 3-16.
- . 1983. "On the Nature of the Psychological." Philosophical Studies 43: 155-164.
- . 1984. "The Primacy of the Intentional." Synthese 61: 89-109.

- Chomsky, Noam. 1978. Rules and Representations. New York: Columbia University Press.
- . 1980. "Crítica de "Verbal Behavior" de B.F. Skinner." ¿Chomsky o Skinner? La génesis del lenguaje. Ed. Ramón Bayés. Barcelona: Fontanella.
- . 1988. Language and Problems of Knowledge. Cambridge, Mass: The MIT Press.
- Churchland, Patricia Smith y Paul M. Churchland. 1981. "Functionalism, Qualia, and Intentionality." Philosophical Topics 12: 121-146.
- Churchland, Paul M. 1981. "Eliminative Materialism and Propositional Attitudes." The Journal of Philosophy 78: 67-90.
- . 1984. Matter and Consciousness: A Contemporary Introduction to the Philosophy of Mind. Cambridge, Mass.: The MIT Press.
- . 1985. "Reduction, Qualia, and the Direct Introspection of Brain States." The Journal of Philosophy 82: 8-28.
- Davidson, Donald. 1979. "Mental Events." Experience and Theory. Ed. Lawrence Foster & J. W. Swanson. Mass.: The University of Massachusetts Press: 79-101.
- . 1985. "La Mente Material." Mentes y Máquinas. Ed. Manuel Garrido. Madrid: Tecnos: 105-126.
- Davis, Lawrence H. 1982. "Functionalism and Absent Qualia." Philosophical Studies 41: 231-250.
- Dennett, Daniel C. 1969. Content and Consciousness. London: Routledge & Kegan Paul.
- . 1978. "Current Issues In the Philosophy of Mind." American Philosophical Quarterly 15: 249-261.
- . 1981. Braimstorms. Cambridge, Mass.: The MIT Press.
- . 1982a. The Intentional Stance. Cambridge, Mass.: The MIT Press.
- . 1982b. "'How to Study Human Consciousness Empirically or Nothing Comes to Mind.'" Synthese 53: 159-180.
- . 1982c. "Comments on Rorty." Synthese 53: 349-356.
- . 1983. "Styles on Mental Representation." Proceedings of the Aristotelian Society 83: 213-226.
- . 1989. "Hacia una teoría cognitiva de la conciencia." Cuadernos de Crítica 46: 5-41.
- Descartes, René. 1967. Oeuvres philosophiques. Tome II (1638-1642). Ed. Ferdinand Alquié. Paris: Editions Garnier.
- . 1977. Meditaciones Metafísicas con Objeciones y Respuestas. Introducción, traducción y notas de Vidal Peña. Madrid: Ediciones Alfaguara. (He seguido la traducción de esta edición.)
- . 1990. El tratado del hombre. Trad. Guillermo Quintás. Madrid: Alianza Editorial.

- Detlefsen, Michael. 1986. Hilbert's Program. An Essay on Mathematical Instrumentalism. Holland: Dordrecht Reidel.
- Donnellan, Keith S. 1972. "Proper Names and Identifying Descriptions." En Semantic and Natural Language. Eds. Harman y Davidson. Dordrecht: Dordrecht Publishing Company: 356-379.
- . 1974. "Speaking of Nothing." Philosophical Review 83: 3-31.
- Donovan, Carol. 1978. "Eliminative Materialism Reconsidered." Canadian Journal of Philosophy 8: 289-303.
- Doore, Gary. 1981. "Functionalism and Absent Qualia." The Australasian Journal of Philosophy 59: 387-402.
- Double, Richard. 1983. "Searle, Programs, and Functionalism." Nature and System 5: 107-114.
- . 1984. "Reply to Fields." Nature and System 6: 55-57.
- Dreske, Fred. 1977. "Causal Theories of Reference." The Journal of Philosophy 74: 621-625.
- . 1988. Explaining Behavior: Reasons in a World of Causes. Cambridge, Mass.: The MIT Press.
- y Berent Enc. 1984. "Causal Theories of Knowledge." Midwest Studies in Philosophy. Causation and Causal Theories. Ed. Peter A. French et al. Minneapolis: University of Minnesota Press: 517-528.
- Dreyfus, Hubert H. 1972. What Computers Can't Do. New York. Harper and Row.
- Dreyfus, Hubert H. ed. 1982. Husserl, Intentionality and Cognitive Science. Cambridge, Mass.: The MIT Press.
- Elugardo, Reinaldo. 1981. "Machine Functionalism and the New Lilliputian Argument." Pacific Philosophical Quarterly 62: 256-261.
- . 1983. "Functionalism and the Absent Qualia Argument." Canadian Journal of Philosophy 13: 161-179.
- Enc, Berent. 1982. "Intentional States of Mechanical Devices." Mind 91: 161-182.
- . 1983. "In Defense of the Identity Theory." The Journal of Philosophy: 279-298.
- . 1989. "Causal Theories and Unusual Causal Pathways." Philosophical Studies 5: 231-261.
- Endicott, Ronald P. 1989. "On Physical Multiple Realization." Pacific Philosophical Quarterly 70: 212-224.
- Eshelman, Larry J. 1977. "Functionalism, Sensations, and Materialism." Canadian Journal of Philosophy 7: 255-274.
- Feldman, Richard. 1986. "Davidson's Theory of Propositional Attitudes." The Canadian Journal of Philosophy 16: 693-712.
- Field, Hartry H. 1978. "Mental Representations." Erkenntnis 13: 9-61.
- . 1986. "Stalnaker on Intentionality: On Robert Stalnaker's "Inquiry"." Pacific Philosophical Quarterly 67: 98-112.

- Fields, C.A. 1984. "Double on Searle's Chinese Room." Nature and System 6: 51-54.
- Finke, R. A. 1979. "The functional equivalence of Mental Images and Errors of Movement." Cognitive Psychology 11: 235-264.
- . 1980. "Levels of Equivalence in Imagery and Perception." Psychological Review 87: 113-132.
- Fischer, John A. 1988. "The Wrong Stuff: Chinese Rooms and the Nature of Understanding." Philosophical Investigations 11: 279-299.
- Fischer, John Martin. 1985. "Functionalism and Propositions." Philosophical Studies 48: 295-311.
- Floistad, Guttorm ed. 1983. Contemporary Philosophy: A New Survey. The Hague: Martinus Nijhoff.
- Fodor, Jerry A. 1975. The Language of Thought. Cambridge, Mass.: Harvard University Press.
- . 1980a. "Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology." The Behavioral and Brain Science 3: 63-109.
- . 1980b. La explicación psicológica: Introducción a la filosofía de la psicología. Madrid: Cátedra.
- . 1981a. Representations. Philosophical Essays on the Foundations of Cognitive Science. Cambridge, Mass.: The MIT Press.
- . 1981b. "Reply to Professor Zaitchik on Physicalism." Philosophy and Phenomenological Research 42: 292-293.
- . 1982. "Cognitive Science and the Twin-Earth Problem." Notre Dame Journal of Formal Logic 23: 98-118.
- . 1984a. El Lenguaje del Pensamiento. Madrid: Alianza Editorial/Alianza Psicología.
- . 1984b. "Semantics, Wisconsin Style." Synthese 59: 231-250.
- . 1985. "Fodor's Guide to Mental Representation: The Intelligent Auntie's Vade-Mecum." Mind 94:76-100.
- . 1986. "Why Paramecia Don't Have Mental Representations." Midwest Studies in Philosophy. Studies in the Philosophy of Mind. Eds. Peter A. French et al. Minneapolis: University of Minnesota Press: 3-23.
- . 1987. Psychosemantics. The Problem of Meaning in the Philosophy of Mind. Cambridge, Mass.: The MIT Press.
- . 1989a. "Review Essays: Remnants of Meaning by Stephen Schiffer." Philosophical and Phenomenological Research 50: 409-423.
- . 1989b. "Making Mind Matter More." Philosophical Topics 17: 59-79.
- . 1990. A Theory of Content and Other Essays. Cambridge, Mass.: The MIT Press.
- Fuentes, J.B. 1989. "¿Funciona de hecho la psicología empírica como una fenomenología del comportamiento?" Introducción a Egon Brunswik: El marco conceptual de la psicología. Madrid: Debate.

- . 1991. "Un caso ejemplar de historia interna en psicología: continuidad entre la crítica conductista radical del conductismo metodológico y la crítica gestaltista del estructuralismo." El Basilisco Segunda Epoca 8: 19-39.
- García-Albea, Juan A. 1991. "La mente como máquina simbólica." Revista de Occidente 119: 47-60.
- . 1991. "Entrevista con Jerry Fodor." Revista de Occidente. 119: 17-46.
- García-Baró López, Miguel. 1983. Fundamentos de la Crítica de la Razón Lógica. Ensayo Fenomenológico. Madrid: Universidad Complutense de Madrid.
- . 1986. "La filosofía primera de Edmund Husserl en torno a 1900." Diánoia: 41-69.
- . Categorías, Intencionalidad y Números. Introducción a la filosofía primera y a los orígenes del pensamiento fenomenológico. (En manuscrito).
- Garfield, Jay L. 1988. Belief in Psychology. A Study in the Ontology of Mind. Cambridge, Mass.: The MIT Press.
- Georgalis, Nicholas. 1986. "Intentionality and Representation." International Studies in Philosophy 18: 45-58.
- Gillet, Grant. 1989. "Representation and Cognitive Science." Inquiry 32: 261-276.
- Goldman, Alvin I. 1977. "Perceptual Objects." Synthese 35: 257-284.
- Graham, George y Stephens G. Lynn. 1985. "Are Qualia a Pain in the Neck for Functionalists?" American Philosophical Quarterly 22: 73-80.
- Gregory, Richard L. 1965. Ojo y Cerebro. Psicología de la Visión. Madrid: Guadarrama.
- Grice, P. 1957. "Meaning." Philosophical Review 66: 377-388.
- . 1969. "Vacuous Names." En Words and Objections. Eds. D. Davidson and J. Hintikka. Dordrecht: Dordrecht Reidel: 118-145.
- Haldane, John J. 1989. "Naturalism and the Problem of Intentionality." Inquiry 32: 305-322.
- Hall Richard J. 1977. "Seeing and Naming." Synthese 35: 381-393.
- Hamlyn, D.W. 1990. In & Out of the Black Box. On the Philosophy of Cognition. Cambridge, Mass.: Basil Blackwell.
- Harman, Gilbert. 1973. Thought. Princeton: Princeton University Press.
- Heil, John. 1981. "Does Cognitive Psychology Rest on a Mistake?" Mind 90: 321-342.
- Hellman, Geoffrey P. y Frank Willson Thompson. 1975. "Physicalism: Ontology, Determination, and Reduction." The Journal of Philosophy 72: 551-564.
- Hempel, Carl G. 1972. "Logical Analysis of Psychology." Intentionality, Mind and Language. Ed. Ausonio Marras. Urbana, Chicago, London: University of Illinois Press: 115-152.

- Hintikka, Jaakko. 1975. "Die Intentionen der Intentionalität." Nue Hefte für Philosophie. 8: 65-95.
- . 1982. "Comments and Replies to Husserl's Phenomenological Conception of Intentionality and Its Difficulties." Philosophia (Israel) 11: 277-286.
- Horgan, Terence. 1981. "Token Physicalism, Supervenience, and the Generality of Physics." Synthese 49: 395-413.
- . 1984. "Functionalism and Token Physicalism." Synthese 59: 321-338.
- Husserl, Edmund. 1950. Ideen zu einer reinen Phänomenologie und phänomenologischen Philosophie. Ed. Walter Biemel. Husserliana III. La Haya: Martinus Nijhoff.
- . 1949. Ideas relativas a una fenomenología pura y a una filosofía fenomenológica. Edición y traducción de José Gaos. México: Fondo de Cultura Económica. (He seguido esta traducción con algunas pequeñas variantes.)
- . 1963. Cartesianische Meditationen und pariser Vorträge. Ed. Stephan Strasser. Husserliana I. La Haya: Martinus Nijhoff.
- . 1985. Meditaciones Cartesianas. Introducción a la fenomenología. Ed. Miguel García-Baró. México: Fondo de Cultura Económica. (He seguido esta traducción con algunas pequeñas variantes.)
- . 1965. Philosophie als strenge Wissenschaft. Frankfurt am Main: Vittorio Klostermann.
- . 1970. Philosophie der Arithmetic. Mit ergänzenden Texten (1890-1901). Ed. Lothar Eley. Husserliana XII. La Haya: Martinus Nijhoff.
- . 1971. Ideen zu einer reinen Phänomenologie und phänomenologischen Philosophie. III. Die Phänomenologie und die Fundamente der Wissenschaften. Ed. Marly Biemel. Husserliana V. The Hague. Martinus Nijhoff.
- . 1976a. Die Krisis der europäischen Wissenschaften und die transzendente Phänomenologie. Ed. Walter Biemel. Husserliana VI. The Hague: Martinus Nijhoff.
- . 1976b. "Naturwissenschaftliche und geisteswissenschaftliche Einstellung. Naturalismus, Dualismus und psychophysische Psychologie." En Die Krisis der europäischen Wissenschaften und die transzendente Phänomenologie. Ed. Walter Biemel. Husserliana VI. The Hague: Martinus Nijhoff.
- . 1984. Logische Untersuchungen. Ed. Ursula Panzer. Husserliana XIX/1, XIX/2. The Hague: Martinus Nijhoff.
- . 1985. Erfahrung und Urteil. Ed. Ludwig Landgrebe. Hamburg: Felix Meiner.
- . 1991. La Crisis de las Ciencias Europeas y la Fenomenología Transcedental. Trad. Jacobo Muñoz y Salvador Mas. Barcelona: Crítica. (He seguido esta traducción con pequeñas variantes.)

- Jackson, Frank, Robert Pargetter y Elizabeth W. Prior. 1982. "Functionalism and Type-Type Identity Theories." Philosophical Studies 42: 209-225.
- Jacquette, Dale. 1989a. "Searle's Intentionality Thesis." Synthese 80: 267-275.
- . 1989b. "Adventures in the Chinese Room." Philosophy and Phenomenological Research 49: 605-623.
- . 1990. "Intentionality and Stich's Theory of Brain Sentence Syntax." The Philosophical Quarterly 40: 169-182.
- Jahren, Neal. 1990. "Can Semantic Be Syntactic?" Synthese 82: 309-328.
- Johson-Laird, P.N. 1977. "Procedural Semantics, 189." Cognition. International Journal of Cognitive Psychology 5: 189-214.
- . 1978. "What's Wrong with Grandma's Guide to Procedural Semantics: A Reply to Jerry Fodor." Cognition. International Journal of Cognitive Psychology 6: 249-261.
- Kern, Iso. 1962. "Die drei Wege zur transzendental-phaenomenologischen Reduktion in der Philosophie Edmund Husserls." Tijdschrift voor Filosofie 24/2: 303-349.
- Kim, Jaegwon. 1977. "Perception And Reference without Causality." The Journal of Philosophy 74: 606-620.
- . 1978. "Supervenience and Nomological Incommensurables." American Philosophical Quarterly 15: 149-156.
- . 1979. "Causality, Identity, and Supervenience in the Mind-Body Problem." Midwest Studies in Philosophy. Studies in Metaphysics. Eds. Peter A. French et al. Minneapolis: University of Minnesota Press: 31-49.
- . 1982. "Psychophysical Supervenience." Philosophical Studies 41: 51-70.
- . 1984. "Epiphenomenal and Supervenient Causation." Midwest Studies in Philosophy. Causation and Causal Theories. Eds. Peter A. French et al. Minneapolis: University of Minnesota Press: 257-270.
- Kitcher, Patricia. 1979. "Phenomenal Qualities." American Philosophical Quarterly 16: 123-129.
- . 1980. "Discussion: How to Reduce a Functional Psychology?" Philosophy of Science 47: 134-140.
- Krantz, Susan. 1990. "Brentano on "Unconscious Consciousness"." Philosophy and Phenomenological Research. 50: 745-753.
- Kripke, Saul A. 1980. Naming and Necessity. Cambridge, Mass.: Harvard University Press.
- . 1988. "A Puzzle about Believe." Propositions and Attitudes. Eds. Nathan Salmon y Scott Soames. New York: Oxford University Press.
- Lackner, J. R y M. Garrett. 1973. "Resolving ambiguity: effects of biasing context in the unattended ear." Cognition. 1: 359-372.
- Lahren, Brian. 1976. "Commentary on Margolis' Paper "Mental States"." Behaviorism 4: 77-95.

- Leibniz, Gottfried Wilhelm. 1960. Nouveaux Essais Sur L'entendement par L'auteur du systeme de l'harmonie preestablie. En Die philosophischen Schriften von Gottfried Wilhelm Leibniz. V. Ed. C.I. Gerhardt. Darmstadt: Georg Olms Hildesheim.
- LePore, Ernest y Barry Loewer. 1989. "More on Making Mind Matter." Philosophical Topics 17: 175-191.
- Lloyd, Dan. 1987. "Mental Representation from the Bottom up." Synthese 70: 23-78.
- Loar, Brian. 1976. "The Semantics of Singular Terms." Philosophical Studies 30: 353-377.
- . 1987. "Subjective Intentionality." Philosophical Topics 15. 1: 89-124.
- Locke, John. 1975. An Essay Concerning Human Understanding. Ed. Peter H. Nidditch. Oxford: Oxford University Press.
- Lockwood, M. 1971. "Identity and Reference." Identity and Individuation. Ed. M. Munitz. New York: New York University Press: 119-211.
- Loewer, Barry. 1987. "From Information to Intentionality." Synthese 70: 287-317.
- Lycan, William G. 1981. "Psychological Laws." Philosophical Topics 12: 9-38.
- . 1990. "Mental Content in Linguistic Form." Philosophical Studies 58: 147-154.
- Lyons, William E. 1985. "On Searle's "Solution" to the Mind-Body Problem." Philosophical Studies 48: 291-294.
- Mackie, J.L. 1979. "Mind, Brain, and Causation." Midwest Studies in Philosophy. Studies in Metaphysics. Eds. Peter A. French et al. Minneapolis: University of Minnesota Press: 19-29.
- Maloney, Christopher J. 1984. "The Mundane Mental Language: How to Do Words with Things." Synthese 59: 251-294.
- . 1987. "The Right Stuff." Synthese 70: 349-372.
- Margolis, Joseph. 1975. "Mental States." Behaviorism 3: 23-31.
- Marras, Ausonio ed. 1972. Intentionality, Mind and Language. Urbana, Chicago, London: University of Illinois Press.
- . 1987. "Intentionality and Probability: Reply to Yoder." Philosophia (Israel) 17: 307-310.
- Martin, Errol B. y Karl Pfeifer. 1986. "Intentionality and the Non-Psychological." Philosophy and Phenomenological Research 46: 531-554.
- Martínez Freire, Pascual. 1993. "Representación e Interpretación." Thémata. Revista de Filosofía 11: 37-46.
- Maudlin, Tim. 1989. "Computation and Consciousness." The Journal of Philosophy 86: 407-432.
- McCann, Hugh J. 1986. "Intrinsic Intentionality." Theory and Decision 20: 247-273.

- McCarthy, Timothy. 1984. "Representation, Intentionality, and Quantifiers." Synthese 60: 369-412.
- McCulloch, Gergory. 1984. "Cause in Perception: A Note on Searle's Intentionality." Analysis 44: 203-205.
- McGinn, Colin. 1980. "Functionalism and Phenomenalism: A Critical Note." The Australasian Journal of Philosophy 58: 35-46.
- . 1981. "A Note on Functionalism and Function." Philosophical Topics 12: 169-170.
- . 1991. The Problem of Consciousness. Essays Towards a Resolution. Cambridge. Mass: Basil Blackwell.
- McIntyre, Ronald. 1982. "Husserl's Phenomenological Conception of Intentionality." Philosophia (Israel) 11: 223-248.
- McLaughlin, Brian P. 1984. "Perception, Causation and Supervenience." Midwest Studies in Philosophy. Causation and Causal Theories. Eds. Peter A. French et al. Minneapolis: University of Minnesota Press: 569-591.
- . 1987. "What is Wrong with Correlational Psychosemantics." Synthese 70: 271-286.
- Merleau-Ponty, Maurice. 1953. La estructura del comportamiento. Buenos Aires: Librería Hachette.
- Mill, Izchak. 1984. "Perceptual Reference." Synthese. 61: 35-59.
- Mohanty, J.N. 1969. Edmund Husserl's Theory of Meaning. The Hague: Martinus Nijhoff.
- . 1972. The Concept of Intentionality. St. Louis: Warren H. Green.
- . 1977. "Husserl's Thesis of the Ideality of Meaning." Readings on Edmund Husserl's "Logical Investigations". Ed. J.N. Mohanty. The Hague: Martinus Nijhoff: 76-82.
- . 1981. "Intentionality and the Noema." The Journal of Philosophy 78: 706-717.
- . 1984. "Intentionality, Causality and Holism." Synthese 61: 17-33.
- y William R. McKenna eds. 1989. Hussel's Phenomenology: A Textbook. Washington, D.C.: Center for Advanced Research in Phenomenology & University Press of America.
- Morris, Katherine J. 1984. "In Defense of Methodological Solipsism: A Reply to Noonan." Phylosophical Studies 45: 399-411.
- Mucciolo, Laurence. 1974. "The Identity Theory and Criteria for the Mental." Philosophy and Phenomenological Reserch 35: 167-180.
- . 1975. "Neurophysiological Reduction, Psychological Explanation and Neuropsychology." Philosophy of the Social Sciences 5: 451-462.
- Nagel, Thomas. 1970. "Physicalism." The Mind-Brain Identity Theory. Ed. C.V. Borst. New York: St. Martin's Press: 214-230.
- Nelson, Alan. 1985. "Physical Properties." Pacific Philosophical Quarterly 66: 268-282.

- Olafson, Frederick A. 1975. "Husserl's Theory of Intentionality in Contemporary Perspective." Nous 9: 73-83.
- Ortega y Gasset, José. 1977. Ideas y creencias y otros ensayos de filosofía. Madrid: Revista de Occidente.
- Otto, Herbert R. y James A. Tuedio eds. 1988. Perspectives on Mind. Dordrecht: Dordrecht Reidel.
- Owens, Joseph. 1983. "Functionalism and Propositional Attitudes." Nous 17: 529-549.
- Papineau, David. 1984. "Representation and Explanation." Phylosophy of Science 51: 550-572.
- . 1990. "Why Supervenience?" Analysis 50: 66-71.
- Peacocke, Christopher. 1979. "Deviant Causal Chains." Midwest Studies in Philosophy. Studies in Metaphysics. Eds. Peter A. French et al. Minneapolis: University of Minnesota Press: 123-155.
- Pears, D. F. 1976. "The Causal Conditions of Perception." Synthese 33: 25-40.
- Pendlebury, Michael. 1987. "Stalnaker on Inquiry." Journal of Philosophical Logic 16: 229-272.
- Platón. 1986. Diálogos. Fedón, Banquete, Fedro. Trad. Carlos García Gual (Fedón). Madrid: Gredos.
- Preston, John M. 1989. "Folk Psychology as Theory or Practice? The Case for Eliminative Materialism." Inquiry 32: 277-303.
- Putnam, Hilary. 1964. "Minds and Machines." Minds and Machines. Ed. Alan Ross Anderson. Englewood Cliffs: Prentice -Hall: 72-97.
- . 1975. "The Meaning of Meaning." Mind, Language and Reality. Philosophical Papers vol. 2. Cambridge: Cambridge University Press.
- . 1981. Reason, Truth and History. Cambridge: Cambridge University Press.
- . 1982a. "Comment on Fodor's "Cognitive Science and the Twin Earth Problem." Notre Dame Journal of Formal Logic 23: 294-295.
- . 1982b. "Why Reason Can't Be Naturalized." Synthese 52: 3-24.
- . 1990. Representación y Realidad. Un balance crítico del funcionalismo. Barcelona: Gedisa.
- Pylyshyn, Zenon W. 1974-75. "Minds, Machines and Phenomenology: Some Reflections on Dreyfus' What Computers can't Do." Cognition 3: 57-77.
- . 1985. Computation and Cognition. Toward a Foundation for Cognitive Science. Cambridge, Mass.: The MIT Press.
- . 1987. "What's in a Mind?" Synthese 70: 97-122.
- Quesada, Daniel. 1992. "Una Imagen de la Filosofía desde la Filosofía de las Ciencias Cognitivas." Quaderns de Filosofia i Ciència. Actes del VIIé Congrés de Filosofia Al País Valencia. 7: 25-38.

- Quine, W. V. 1974. "Objetos Proposicionales." La relatividad ontológica y otros ensayos. Madrid: Tecnos: 177-201.
- . 1985. "States of Mind." The Journal of Philosophy 82: 5-8.
- Quinton, Anthony. 1955. "The Problem of Perception." Mind. 64: 28:51.
- . 1973. The Nature of Things. Londres: Routledge and Keagan Paul.
- Rang, Bernhard. 1973. Kausalität und Motivation. The Hague: Martinus Nijhoff.
- Rapaport, William J. 1986. "Discussion: Searle's Experiment with Thought." Philosophy of Science 53: 271-279.
- Reed, Edward S. 1983. "Two Theories of the Intentionality of Perceiving." Synthese 54: 85-94.
- Rey, George. 1986. "What's Really Going on in Searle's "Chinese Room". " Philosophical Studies 50: 169-185.
- Richard, Mark. 1990. Propositional Attitudes. An Essay on Thought and How We Ascribe Them. Cambridge: Cambridge University Press.
- Richardson, Robert C. 1981. "Internal Representation: Prologue to a Theory of Intentionality." Philosophical Topics 12: 171-211.
- Rock, Irvin. 1985. La Percepción. Barcelona: Labor
- Rorty, Richard. 1970. "Mind-Body Identity, Privacy and Categories." The Mind-Brain Identity Theory. Ed. C.V. Borst. New York: St. Martin's Press: 187-213.
- . 1982a. "Contemporary Philosophy of Mind." Synthese 53: 323-348.
- . 1982b. "Comments on Dennett." Synthese 53: 181-187.
- Rosenthal, David M. 1986. "Intentionality." Midwest Studies in Philosophy. Studies in the Philosophy of Mind. Eds. Peter A. Frenck et al. Minneapolis: University of Minnesota Press: 3-23.
- Russell, Bertrand. 1956. "Knowledge by Acquaintance and Knowledge by Description." En Mysticism and Logic and other Essays. New York: Barnes & Noble.
- . 1984. "On Denoting." En Logic and Knowledge. Essays 1901-1950. Ed. Robert Charles Marsh. London: George Allen and Unwin: 41-56.
- . 1989. The Problems of Philosophy. Oxford: Oxford University Press.
- Salmon, Nathan. 1981. Reference and Essence. Princeton N.J.: Princeton University Press.
- Salmon, Nathan y Scott Soames eds. 1988. Propositions and Attitudes. New York: Oxford University Press.
- Sanfélix Vidarte, Vicente. 1991. "Panorama actual de la filosofía analítica de la mente." En Filosofía analítica hoy. Encuentro de tradiciones. Mercedes Torrevejano (Coord.). Santiago de Compostela: Universidad de Santiago de Compostela: 121-149.
- San Martín, Javier. 1992. "Psicología Cognitiva y Fenomenología." Quaderns de Filosofia i Ciència. Actes del VIIé Congrés de Filosofia Al País Valencia. 7: 7-24.

- Sayre, Ken. 1987. "Cognitive Science and the Problem of Semantic Content." Synthese 70: 247-269.
- Schiffer, Stephen. 1982. "Intention-Based Semantics." Notre Dame Journal of Formal Logic 23: 119-156.
- . 1986. "Stalnaker's Problem of Intentionality: On Robert Stalnaker's 'Inquiry'." Pacific Philosophical Quarterly 67: 87-97.
- . 1989. Remnants of Meaning. Cambridge, Mass.: The MIT Press.
- Seager, W. 1983. "Functionalism, Qualia and Causation." Mind 92: 174-188.
- Searle, John R. 1979. "What is an Intentional State?" Mind 88: 74-92.
- . 1980. "Minds, Brains and Programs." The Behavioral and Brain Sciences 3: 417-424.
- . 1981a. "Analytic Philosophy and Mental Phenomena." Midwest Studies in Philosophy. The Foundations of Analytic Philosophy. Eds. Peter A. French et al. Minneapolis: University of Minnesota Press: 405-423.
- . 1981b. "Intentionality and Method." The Journal of Philosophy 78: 720-733.
- . 1982. "Proper Names and Intentionality." Pacific Philosophical Quarterly 63: 205-225.
- . 1983. Intentionality. An Essay in the Philosophy of Mind. Cambridge: Cambridge University Press.
- . 1984a. "Intentionality and Its Place in Nature." Synthese 61: 3-16.
- . 1984b. Minds, Brains and Science. Cambridge, Mass.: Harvard University Press.
- . 1989a. "Reply to Jacqueline." Philosophy and Phenomenological Research 49: 701-708.
- . 1989b. "Consciousness, Unconsciousness, and Intentionality." Philosophical Topics 17: 193-209.
- Sebastián y Zúñiga, José María. Introducción a la Visión Artificial. Madrid: División de Ingeniería de Sistemas y Automática (DISAM) de la Universidad Politécnica de Madrid.
- Seidel, Asher. 1988. "Searle on the Biological Basis of Cognition." Analysis: 26-28.
- . 1989. "Chinese Room A, B, and C." Pacific Philosophical Quarterly 70: 167-173.
- Sellars, Wilfrid. 1963. "Philosophy and the Scientific Image of the Man." En Science, Perception and Reality. London: Routledge & Kegan Paul: 1-40.
- . 1972a. "Empiricism and the Philosophy of Mind." Intentionality, Mind and Language. Ed. Ausonio Marras. Urbana, Chicago, London: University of Illinois Press: 197-213.
- . 1972b. "Notes on Intentionality." Intentionality, Mind and Language. Ed. Ausonio Marras. Urbana, Chicago, London: University of Illinois Press: 321-334.

- Serrano de Haro, Agustín. 1990. Fenomenología Transcendental y Ontología. Madrid: Universidad Complutense de Madrid.
- Sher, George. 1975. "Sentences in the Brain?" Philosophy and Phenomenological Research 36: 94-99.
- Shoemaker, Sydney. 1975. "Functionalism and Qualia." Philosophical Studies 27: 291-315.
- . 1981. "Some Varieties of Functionalism." Philosophical Topics 12: 93-119.
- Silvers, Stuart. 1989. "Representational Capacity, Intentional Ascriptions, and the Slippery Slope." Philosophy of Science 56: 463-473.
- Smith, David Woodruff, y Ronald McIntyre. 1982. Husserl and Intentionality. A Study of Mind, Meaning, and Language. Boston: Dordrecht Reidel.
- Smith, David Woodruff. 1983. "Husserl's Philosophy of Mind." Contemporary Philosophy: A New Survey. Ed. Guttorm Floistad. The Hague: Martinus Nijhoff: 249-286.
- . 1984. "Content and Context of Perception." Synthese 61: 61-87.
- Sobert, Elliot. 1976. "Mental Representations." Synthese 33: 101-148.
- . 1980. "Language and Psychological Reality." Linguistics and Philosophy 3: 395-405.
- Solomon, Robert C. 1976. "Psychological Predicates." Philosophy and Phenomenological Research 36: 472-493.
- Sosa, Ernest. 1984. "Mind-Body Interaction and Supervenient Causation." Midwest Studies in Philosophy. Causation and Causal Theories. Eds. Peter A. French et al. Minneapolis: University of Minnesota Press: 271-281.
- Stalnaker, Robert. 1984. Inquiry. Cambridge, Mass.: The MIT Press.
- . 1986. "Replies to Schiffer 'Stalnaker's Problem of Intentionality' and Field's 'Stalnaker on Intentionality'." Pacific Philosophical Quarterly 67: 113-123.
- . 1987. "Semantics for Belief." Philosophical Topics 15.1: 177-190.
- . 1990. "Mental Content and Linguistic Form." Philosophical Studies 58: 129-146.
- Stampe, Dennis W. 1977. "Toward a Causal Theory of Linguistic Representations." Midwest Studies in Philosophy. Studies in the Philosophy of Language. Eds. Peter A. French et al. Minneapolis: University of Minnesota Press: 42-63.
- . 1986. "Verificationism and a Causal Account of Meaning." Synthese 69: 107-137.
- Stevelný, Kim. 1989. "Fodor's Nativism." Philosophical Studies 55: 119-141.
- Stich, Stephen. 1981. "Dennett on Intentional Systems." Philosophical Topics 12: 39-62.

- . 1983. From Folk Psychology to Cognitive Science: The Case Against Belief. Cambridge, Mass.: The MIT Press.
- . 1990. "Building Belief: Some Queries about Representation, Indication, and Function." Philosophy and Phenomenological Research. 50: 801-806.
- Strawson, P.F. 1985. "On Referring". The Philosophy of Language. Ed. A.P. Martinich. Oxford: Oxford University Press: 220-235.
- Taylor, Kenneth A. 1987. "Belief, Information and Semantic Content: A Naturalist's Lament." Synthese 71: 97-124.
- . 1989. "Narrow Content Functionalism and the Mind-Body Problem." Nous 23: 355-372.
- Terrell, Burnham. 1983. "Brentano's Philosophy of Mind." Contemporary Philosophy: A New Survey. Ed. Guttorm Floistad. The Hague: Martinus Nijhoff: 223-247.
- Thompson, David L. 1986. "Intentionality and Causality in John Searle." Canadian Journal of Philosophy 16: 83-97.
- Tichy, Mattias. 1987. "Auf Der Suche Nach Dem Subjekt In Der Maschine: Zu John R. Searles Kritik Am Computermodeil Des Denkens." Zeitschrift für Philosophische Forschung 41: 429-443.
- Tichy, Pavel. 1986. "Putnam on Brain in the Vat." Philosophia (Israel) 16: 137-146.
- Toberlin, James. 1984. "Identity, Intensionality and Intentionality." Synthese 61: 111-131.
- Turing, Alan A. 1950. "Computing Machinery and Intelligence." Mind 59: 433-460.
- Twardowski, Kasimir. 1982. Zur Lehre vom Inhalt und Gegenstand der Vorstellungen. Eine psychologische Untersuchung. Munich: Philosophia.
- Tye, Michael. 1983. "Functionalism and Type Physicalism." Philosophical Studies 44: 161-174.
- Van Gulick, Robert. 1980. "Functionalism, Information, and Content." Natur and System 2: 139-162.
- . 1982. "Mental Representation. A Functionalist View." Pacific Philosophical Quarterly 63: 3-20.
- Vega, Manuel de. 1984. Introducción a la Psicología Cognitiva. Madrid: Alianza Editorial/Alianza Psicología.
- White, Nicholas P. 1985. "Professor Shoemaker and So-Colled 'Qualia' of Experience." Philosophical Studies 47: 369-383.
- White, Stephen L. 1987. "What Is It Like to Be an Homunculus?" Philosophical Quarterly 68: 148-174.
- Wider, Kathleen. 1990. "Overtones of Solipsism in Thomas Nagel's "What Is Like to Be a Bat?" and the View from Nowhere." Philosophy and Phenomenological Research 50: 481-499.

- Wilkes, K.V. 1981. "Functionalism, Psychology, and the Philosophy of Mind." Philosophical Topics 12: 147-167.
- Wilks, Yorick. 1990. "Form and Content in Semantics." Synthese 82: 329-351.
- Wilson, George. 1979. "Cheap Materialism." Midwest Studies in Philosophy. Studies in Metaphysics. Eds. Peter A. French et al. Minneapolis: University of Minnesota Press: 51-72.
- Winograd, Terry. 1991. "Máquinas pensantes: ¿son posibles?, ¿lo somos?" Revista de Occidente 119: 113-150.
- Wittgenstein, Ludwig. 1984. Los cuadernos azul y marrón. Madrid: Tecnos.
- Wright, Larry. 1973. "Functions." The Philosophical Review 82: 139-168.
- Yoder, Jesse L. 1987. "Chisholm's Criteria of Intentionality." Philosophia (Israel) 17: 297-305.
- Zaitchik, Alan. 1981a. "Intentionalism and Physical Reductionism in Computational Psychology." Philosophy and Phenomenological Research 42: 23-41.
- . 1981b. "Reply to Professor Fodor on Physicalism." Philosophy and Phenomenological Research 42: 294-295.